



FACULTAD DE INFORMÁTICA
UNIVERSIDAD POLITÉCNICA DE MADRID

TÉSIS DE MÁSTER
MÁSTER UNIVERSITARIO EN
INTELIGENCIA ARTIFICIAL

**CLUSTERING BASADO EN
REDES BAYESIANAS CON
PREDICTORAS CONTINUAS.
APLICACIONES EN
NEUROCIENCIA**

AUTOR : Sergio Luengo Sánchez
TUTORES : Concha Bielza Lozoya
Pedro Larrañaga Mugica

Nov, 2014

Reconocimientos

Reconocer la labor personal y académica de todas las personas que de una manera u otra han aportado algo a este proyecto, y hacerlo en pocas líneas, no hace honor a lo que han contribuido.

En primer lugar quiero agradecer a mis mentores Concha Bielza y Pedro Larrañaga su confianza y su apoyo, así como el conocimiento y experiencia que me han transmitido que han sido fundamentales para la realización de este proyecto. Ha sido un privilegio desarrollar este trabajo bajo su tutela.

También quiero reconocer la labor de Javier de Felipe, Ruth Benavides-Piccione e Isabel Fernaud Espinosa del Laboratorio Cajal de Circuitos Corticales(CSIC) por facilitarme la inmersión en la investigación interdisciplinar.

Este trabajo no hubiese sido posible sin el apoyo financiero del proyecto Cajal Blue Brain (C080020-09).

También quiero agradecer a mis compañeros del Computational Intelligence Group (CIG) su afabilidad desde el primer día así como toda la ayuda que me han prestado durante todo el proceso.

Finalmente quiero agradecer a mi familia y amigos por su compañía y aliento. Mi mayor gratitud va dirigida a Antonio, Encarna, Sara y Ana por su amor y ánimos. Este trabajo está dedicado a ellos.

Resumen

Desentrañar el funcionamiento del cerebro es uno de los principales desafíos a los que se enfrenta la ciencia actual. Un área de estudio que ha despertado muchas expectativas e interés es el análisis de la estructura cortical desde el punto de vista morfológico, de manera que se cree una simulación del cerebro a nivel molecular. Con ello se espera poder profundizar en el estudio de numerosas enfermedades neurológicas y patológicas. Con el desarrollo de este proyecto se persigue el estudio del soma y de las espinas desde el punto de vista de la neuromorfología teórica.

Es común en el estado del arte que en el análisis de las características morfológicas de una neurona en tres dimensiones el soma sea ignorado o, en el mejor de los casos, que sea sustituido por una simple esfera. De hecho, el concepto de soma resulta abstracto porque no se dispone de una definición estricta y robusta que especifique exactamente donde finaliza y comienzan las dendritas. En este proyecto se alcanza por primera vez una definición matemática de soma para determinar qué es el soma. Con el fin de simular somas se ahonda en los atributos utilizados en el estado del arte. Estas propiedades, de índole genérica, no especifican una morfología única. Es por ello que se propone un método que agrupe propiedades locales y globales de la morfología. En disposición de las características se procede con la categorización del cuerpo celular en distintas clases a partir de un nuevo subtipo de red bayesiana dinámica adaptada al espacio. Con ello se discute la existencia de distintas clases de somas y se descubren las diferencias entre los somas piramidales de distintas capas del cerebro. A partir del modelo matemático se simulan por primera vez somas virtuales.

Algunas morfologías de espinas han sido atribuidas a ciertos comportamientos cognitivos. Por ello resulta de interés dictaminar las clases existentes y relacionarlas con funciones de la actividad cerebral. La clasificación más extendida (Peters y Kaiserman-Abramof, 1970) presenta una definición ambigua y subjetiva dependiente de la interpretación de cada individuo y por tanto discutible. Este estudio se sustenta en un conjunto de descriptores extraídos mediante una técnica de análisis topológico local para representaciones 3D. Sobre estos datos se trata de alcanzar el conjunto de clases más adecuado en el que agrupar las espinas así como de describir cada grupo mediante reglas unívocas. A partir de los resultados, se discute la existencia de un continuo de espinas y las propiedades que caracterizan a cada subtipo de espina.

Abstract

Unravel how the brain works is one of the main challenges faced by current science. A field of study which has aroused great expectations and interest is the analysis of the cortical structure from a morphological point of view, so that a molecular level simulation of the brain is achieved. This is expected to deepen the study of many neurological and pathological diseases. This project seeks the study of the soma and spines from the theoretical neuromorphology point of view.

In the state of the art it is common that when it comes to analyze the morphological characteristics of a three dimension neuron the soma is ignored or, in the best case, it is replaced by a simple sphere. In fact, the concept of soma is abstract because there is not a robust and strict definition on exactly where it ends and dendrites begin. In this project a mathematical definition is reached for the first time to determine what a soma is. With the aim to simulate somas the attributes applied in the state of the art are studied. These properties, generic in nature, do not specify a unique morphology. It is why it was proposed a method to group local and global morphology properties. In arrangement of the characteristics it was proceed with the categorization of the celular body into different classes by using a new subtype of dynamic Bayesian network adapted to space. From the result the existance of different classes of somas and differences among pyramidal somas from distinct brain layers are discovered. From the mathematical model virtual somas were simulated for the first time.

Some morphologies of spines have been attributed to certain cognitive behaviours. For this reason it is interesting to rule the existent classes and to relate them with their functions in the brain activity. The most extended classification (Peters y Kaiserman-Abramof, 1970) presents an ambiguous and subjective definition that relies on the interpretation of each individual and consequently it is arguable. This study was based on the set of descriptors extracted from a local topological analysis technique for 3D representations. On these data it was tried to reach the most suitable set of classes to group the spines as well as to describe each cluster by unambiguous rules. From these results, the existance of a continuum of spines and the properties that characterize each spine subtype were discussed .

Contenidos

Lista de imágenes	IX
Lista de tablas	XI
1. Introducción	1
1.1. Neuronas piramidales	3
1.1.1. Soma	3
1.1.2. Dendritas	4
1.1.3. Espinas	4
1.2. Objetivos	6
1.3. Planificación de la tesis	7
 I Soma	 9
2. Extracción de los datos y pre-procesamiento de la neurona	11
2.1. Extracción de los datos	11
2.2. Reparación del soma	12
2.2.1. Reconstrucción de la malla	15
2.3. Segmentación automática de la neurona	16
2.3.1. Bondad del método de segmentación del soma	20
 3. Caracterización del soma	 23
3.1. Caracterización basada en el estado del arte	23
3.1.1. Descriptores globales	24
3.1.2. Descriptores basados en similaridad morfológica	26
3.2. Caracterización basada en Reeb graph	27
3.2.1. Ray-tracer graph	28
3.2.2. Descriptores del soma	30

CONTENIDOS

4. Modelización del soma	33
4.1. Definición del modelo	33
4.1.1. Red bayesiana gaussiana lineal	34
4.2. Aprendizaje de la estructura de la red	36
4.2.1. Red bayesiana dinámica espacial	37
4.2.2. Procedimiento bootstrap para aprendizaje robusto	38
4.3. Clustering probabilístico de somas basado en una red bayesiana dinámica espacial	41
4.3.1. Algoritmo Esperanza-Maximización	41
4.3.1.1. Paso de esperanza	41
4.3.1.2. Paso de maximización	42
4.3.2. Selección del número de clases	44
4.4. Resultados	45
4.4.1. Relaciones de dependencia entre los descriptores	45
4.4.1.1. Dependencia inter-sección	45
4.4.1.2. Dependencia entre secciones	46
4.4.2. Distribución de las clases	47
4.4.2.1. Respecto a la localización cerebral	47
4.4.2.2. Respecto a la similitud morfológica	47
5. Simulación de somas e interpretación del modelo	51
5.1. Simulación	51
5.1.1. Muestro de somas de cada clase	51
5.1.2. Visualización 3D	52
5.1.2.1. Cálculo de las curvas entre secciones	53
5.1.2.2. Generación y reparación de las caras de las secciones	55
5.2. Interpretación del modelo	56
5.2.1. Características segregadoras	56
5.2.1.1. Descriptores del Ray-tracer graph	56
5.2.1.2. Entre áreas	58
5.2.1.3. Variables emergentes	60
5.2.2. Clasificación del soma de acuerdo a la similitud morfológica	61

II	Espinas	63
6.	Obtención, preprocesamiento y caracterización de las espinas	65
6.1.	Introducción	65
6.2.	Extracción de los datos	66
6.2.1.	Reparación de las espinas	67
6.3.	Caracterización basada en Reeb graph	68
6.4.	Descriptores de las espinas	70
7.	Modelización de las espinas	75
7.1.	Definición del modelo	75
7.1.1.	Clustering probabilístico de espinas basado en mixtura de gaussianas	76
7.2.	Distribución de las espinas por cluster	77
8.	Interpretación del modelo	81
8.1.	Caracterización de las clases	81
8.1.1.	Algoritmo RIPPER	81
8.1.2.	Características segregadoras	83
8.1.2.1.	Descripción de las espinas con 7 clases	84
8.1.2.2.	Descripción de las espinas con 4 clases	85
8.2.	Visualización de las clusters	86
8.2.1.	Escalamiento multidimensional	86
8.2.2.	Proyección de las espinas	87
8.3.	Distribución de las espinas por clases en las dendritas	90
8.3.1.	Distribución a partir del compartimento dendrítico, la edad y su combinación	90
8.3.2.	Distribución de las espinas a partir de la distancia al soma	93
9.	Conclusiones y trabajo futuro	99
9.1.	Conclusiones	99
9.1.1.	Somas	99
9.1.2.	Espinas	100
9.2.	Trabajo futuro	101
	Apéndices	105
	Resumen de las características de las espinas	107

CONTENIDOS

Reglas generadas con RIPPER	111
Bibliografía	115

Lista de imágenes

1.1. Dibujo esquemático de una neurona	2
1.2. Neurona piramidal	3
1.3. Clases de espinas	5
1.4. Espinas a lo largo de la dendrita	6
1.5. Fases del análisis de datos	8
2.1. Tinción del soma	12
2.2. Estado inicial de los somas	12
2.3. Proceso de reparación y segmentación	13
2.4. Ejemplo de oclusión ambiental en 2D	13
2.5. Oclusión ambiental aplicada a las neuronas	14
2.6. Neuronas agujereadas después de eliminar los vértices por encima del umbral	15
2.7. Reconstrucción para mallas	16
2.8. Reconstrucción de las neuronas	17
2.9. Ejemplo de <i>Shape Diameter Function</i>	17
2.10. SDF aplicado a neuronas	18
2.11. Histograma y primer clustering	19
2.12. Neuronas tras el primer clustering	19
2.13. Histograma y segundo clustering	20
2.14. Resultado final	20
2.15. Bondad del método de segmentación del soma	21
3.1. Ray-tracer graph	29
3.2. Curvas próximas a elipses	30
4.1. Red bayesiana dinámica espacial	35
4.2. Estructura de la red bayesiana en las secciones i e $i + 1$	40

LISTA DE IMÁGENES

4.3. Selección del número de clusters	45
4.4. Distribución de los somas de acuerdo a la localización cerebral de la que se extrajeron	48
4.5. Distribución de los somas de acuerdo a la similitud morfológica	49
5.1. Proceso de visualización 3D	53
5.2. Triangulación de los puntos de las curvas	56
5.3. Generación de las caras	56
5.4. Ejemplos de somas simulados	57
5.5. Clasificación de somas de acuerdo al área cerebral y la similaridad morfológica	62
6.1. Extracción de espinas	69
6.2. Modelo geométrico particionado y construcción del Reeb graph	71
6.3. Ejemplo de espina extraída y reconstruida	72
6.4. Descripción de las características de las espinas	74
7.1. Valores BIC de acuerdo al número de clases	78
8.1. Ejemplo de espinas para cada una de las 7 clases	85
8.2. Ejemplo de espinas para cada una de las 4 clases	86
8.3. Proyección 3D de la distribución 7-dimensional	88
8.4. Proyección 2D de la distribución 4-dimensional	89
8.5. Distribución de las espinas de acuerdo a p^* para 7 clases	91
8.6. Distribución de las espinas de acuerdo a p^* para 4 clases	93
8.7. Distribución de las 7 clases de acuerdo a la distancia al soma	96
8.8. Distribución de las 4 clusters de acuerdo a la distancia al soma	97
9.1. Red bayesiana dinámica espacial híbrida	102

Lista de tablas

2.1.	Diferencias en la segmentación del soma de expertos y del algoritmo .	21
2.2.	Diferencias en la segmentación del soma del experto 1 sobre el mismo soma distintos días	22
2.3.	Diferencias en la segmentación del soma del experto 2 sobre el mismo soma distintos días	22
4.1.	Conjunto de datos para el aprendizaje de la estructura de la red . . .	38
5.1.	Características discriminatorias entre clases	58
5.2.	Características discriminatorias entre áreas	59
5.3.	Características discriminatorias globales entre áreas	61
6.1.	Distribución de las espinas	67
7.1.	Análisis de la continuidad de espinas con 7 clases	79
7.2.	Análisis de la continuidad de espinas con 4 clases	80
8.1.	Variables discriminatorias de acuerdo al algoritmo RIPPER para 7 clusters	83
8.2.	Variables discriminatorias de acuerdo al algoritmo RIPPER para 4 clusters	84
8.3.	Test de hipótesis entre cada clase y los compartimentos dendríticos,la edad y la combinación de ambas para 7 clusters	92
8.4.	Test de hipótesis entre cada clase y los compartimentos dendríticos,la edad y la combinación de ambas para 4 clases	94
8.5.	Test de hipótesis entre cada clase y la distancia al soma para 7 clases	95
8.6.	Test de hipótesis entre cada clase y la distancia al soma para 4 clases	95
1.	Resumen estadístico de la altura de cada sección.	107
2.	Resumen estadístico del área de cada sección.	107

LISTA DE TABLAS

3. Resumen estadístico del achatamiento y elongación entre cada par de secciones. 108
4. Resumen estadístico del ratio entre los ejes de cada par de secciones. 108
5. Resumen estadístico del coseno del ángulo entre cada par de secciones. 108
6. Resumen estadístico del ángulo colatitud entre cada par de secciones.
La primera sección no fue incluida porque es un valor constante. . . . 109
7. Resumen estadístico del ángulo azimutal entre cada par de secciones. 109

1

Introducción

Desentrañar el funcionamiento del cerebro es uno de los principales desafíos a los que se enfrenta la ciencia actual. Con su estudio, los neurocientíficos pretenden resolver las incógnitas y misterios que lo han acompañado desde hace siglos.

El estudio del cerebro tiene su origen en la Antigüedad, en la que los médicos pensaban que estaba compuesto de “flema”. Posteriormente, Aristóteles estableció que el cerebro era un refrigerador que contrarrestaba el calor del corazón (Zimmer y Clark, 2014). El enigma acerca del encéfalo continuó a lo largo del Renacimiento en el que los anatomistas indicaban que las percepciones, emociones, razonamientos y acciones eran el resultado de “espíritus animales”, vapores místicos e intangibles que se arremolinaban en las cavidades de nuestras cabezas y viajaban por nuestro cuerpo. No es hasta a partir del siglo XVII cuando gracias a Willis (1663) se produjo una revolución debido a que se relacionaron los tejidos cerebrales con la idea de mundo mental.

Tuvo que pasar un siglo para que se comprendiera que el cerebro es un órgano eléctrico. A pesar de ello, poco o nada se sabía acerca de las rutas seguidas por las conexiones del sistema nervioso. Fue Golgi el que introdujo la idea de una red de conexiones sin interrupciones, trabajo que recogió Ramón y Cajal y que amplió aplicando nuevos métodos de tinción, lo que le permitió observar que cada neurona es una célula distinta separada de todas las demás dando lugar a lo que se conoce como la *doctrina de la neurona*. Además, descubrió que las neuronas envían señales a través de unas prolongaciones llamadas axones y que las reciben a través de las prolongaciones receptoras denominadas dendritas (Figura 1.1). Ramón y Cajal, junto con otros científicos, continuó su trabajo analizando la variedad de patrones neuronales y realizando hipótesis acerca de los roles que desempeñaban ciertas morfologías

1. INTRODUCCIÓN

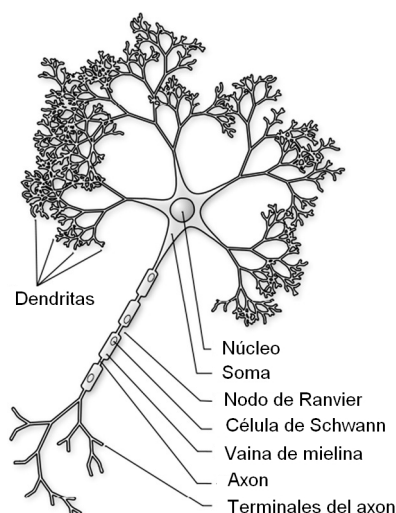


Figura 1.1: Dibujo esquemático de una neurona: En la imagen se muestran los componentes principales de una neurona. Se pueden observar el soma o cuerpo celular donde se encuentra el núcleo de la neurona, el axon o neurita por donde se transmite un impulso nervioso desde el soma hacia otra neurona y las dendritas que son ramificaciones que surgen a partir del núcleo y cuya función principal es la recepción de estímulos. Fuente: Stiles y Jernigan (2010)

de acuerdo a las posiciones cerebrales que les correspondían. Con ello comenzó el estudio de la denominada neuromorfología, esto es, el estudio de la forma y estructura del sistema nervioso.

El trabajo científico anterior supuso la base sobre la que se fundamentan los estudios actuales. Todo ello ha dado lugar a objetivos más ambiciosos relacionados con la comprensión del cerebro y que involucran a multitud de equipos científicos interdisciplinarios. En concreto, un área de estudio que ha despertado muchas expectativas e interés es el análisis de la estructura cortical desde el punto de vista morfológico del cerebro de mamíferos, de manera que se pueda crear una simulación de todo el cerebro a nivel molecular. Con el desarrollo de este tipo de proyectos se pretende comprender mejor su funcionamiento y de esta manera poder profundizar en el estudio de numerosas enfermedades neurológicas y patológicas. Es en este punto en el que la neuromorfología juega un papel fundamental y en la que una de sus ramas, la neuromorfología teórica, es aplicada como medio para alcanzar tales fines. Con la introducción de la neuromorfología teórica se pretende describir la forma y la conectividad del sistema nervioso. De esta manera se han impuesto dos ramas fundamentales. Por un lado se encuentran aquellos científicos que tratan de representar el

cerebro a partir de su actividad sináptica entendiendo por sinapsis la estructura que permite la transducción de señales de una neurona a una célula objetivo. Por otro lado, los que analizan las características estructurales de componentes concretos, es decir, un análisis individualizado neurona por neurona de características morfológicas que permitan asociar las formas de los componentes a los roles que desempeñan. El desarrollo de este proyecto se basará en el segundo método de investigación y más concretamente se centrará en componentes de una neurona piramidal.

1.1. Neuronas piramidales

Las neuronas piramidales (Figura 1.2) fueron descubiertas y estudiadas por Ramón y Cajal que las denominó así debido a la forma piramidal del soma. Este tipo de neuronas son abundantes en la corteza cerebral de prácticamente todos los mamíferos que han sido estudiados, así como en la de los pájaros, peces y reptiles. Esto indica que su existencia en el sistema nervioso tiene un valor adaptativo para el organismo y que sus funciones básicas se han conservado, incluso a medida que evolucionaban, para asumir funciones especializadas. Las neuronas piramidales se encuentran en la mayoría de las estructuras del prosencéfalo, incluyendo la corteza cerebral, el hipocampo y el cuerpo amigdalino. Por lo tanto, las neuronas piramidales se encuentran sobre todo en estructuras que están asociadas con funciones cognitivas avanzadas. El entendimiento de estas neuronas es necesario para dilucidar las bases neuronales de las funciones más sofisticadas.

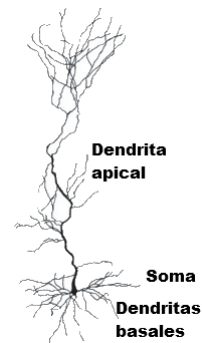


Figura 1.2: Neurona piramidal. Fuente: Spruston (2008)

1.1.1. Soma

El soma, también conocido como cuerpo celular o perikaryon es la región de la neurona que contiene el núcleo. Dado que se trata del cuerpo celular, contiene los típicos orgánulos de las células vivas (mitocondrias, aparato de Golgi, ribosomas, lisosomas,...) lo que convierte al soma en el centro metabólico de la neurona. A partir de estos componentes se realiza el procesamiento químico que da lugar a los neurotransmisores. También se encarga de producir las proteínas para las dendritas, axones y terminales sinápticas.

1. INTRODUCCIÓN

El tamaño del soma de una neurona puede oscilar entre los 0.005mm y los 0.1mm en los mamíferos. Su aspecto es muy variable tomando diversas formas siendo las más representativas la de estrella, la fusiforme, la cónica, la poliédrica, la esférica y la piramidal. En concreto, la piramidal suele representarse como un tetraedro con el ángulo agudo apuntando hacia la superficie del cortex.

1.1.2. Dendritas

Las dendritas son extensiones del cuerpo celular de la neurona. Su función principal consiste en recibir y procesar las señales sinápticas de entrada. Las dendritas muestran gran diversidad estructural pero conociendo la forma del árbol dendrítico se puede indicar el tipo de conectividad que existe entre ciertas neuronas.

El árbol dendrítico de una neurona piramidal se divide en dos tipos, las dendritas basales que surgen de la base del soma emanando una arborización con forma esférica y la dendrita apical que lo hace desde el vértice del soma. Todas las neuronas piramidales tienen varias dendritas basales que suelen ser relativamente cortas. Usualmente, una dendrita apical de gran longitud conecta el soma a un manojo de dendritas. Las características de las dendritas de una neurona piramidal pueden variar considerablemente entre distintas capas, regiones corticales y especies.

1.1.3. Espinas

Las dendritas están cubiertas con miles de espinas dendríticas, esto es, pequeñas protuberancias membranosas cada una de las cuales es receptora de una sinapsis excitatoria (Nimchinsky et al., 2002). Aunque su tamaño y forma son bastante heterogéneos, todas se componen de una cabeza conectada mediante un fino cuello al tronco dendrítico. Las espinas pueden encontrarse en muchas especies, desde anélidos a primates, y son especialmente abundantes en el sistema nervioso central de los vertebrados. Su predominio indica que deben de ser esenciales para el funcionamiento del cerebro. De hecho, en la mayoría de las áreas del cerebro, las espinas son los elementos estructurales dominantes que cubren las dendritas de las neuronas principales. Por ejemplo, las espinas pueden existir en un gran número, incluso más de 200.000 espinas por neurona en las células cerebrales de Purkinje, y son también extraordinariamente abundantes en las dendritas de las neuronas piramidales en el cortex.

Al parecer las espinas tratan de extender la superficie de la membrana dendrítica habilitando los contactos sinápticos, esto es, realizan un trabajo similar al del vello intestinal que incrementa la superficie del área de absorción en el aparato digestivo. De hecho se sabe que las espinas reciben la mayoría de entradas excitatorias y, además, que prácticamente todas las espinas tienen una sinapsis excitatoria en su cabeza. Esto hace pensar que cada espina se corresponde esencialmente con una sinapsis excitatoria.

El número de espinas representa una estimación mínima del número de entradas sinápticas excitatorias en una neurona, que varía ostensiblemente dependiendo de las regiones y especies. Las espinas son comúnmente pequeñas, no llegando a alcanzar los $3\text{ }\mu\text{m}$ de longitud, con una cabeza aproximadamente esférica de entre 0.5 y $1.5\text{ }\mu\text{m}$ de diámetro que se conecta a la dendrita a partir de un cuello estrecho de menos de $0.5\text{ }\mu\text{m}$ de diámetro (Smith et al., 2007). Debido a su tamaño, una dendrita puede albergar más de 50 espinas en menos de $10\text{ }\mu\text{m}$.

Las espinas muestran gran diversidad de morfologías. Se han propuesto muchos tipos de clasificaciones de acuerdo a sus características morfológicas, aunque también se ha discutido si esta diversidad es producto de un continuo en el que no se pueden distinguir claramente las distintas clases (Herrera-Arellano et al., 2007). Sin embargo, quizá la clasificación más exten-

dida es la propuesta por Peters y Kaiserman-Abramof (1970) (Figura 1.3). Las espinas conocidas como *Thin* son las más comunes compuestas por un cuello fino y alargado y una pequeña cabeza bulbosa. Las espinas denominadas *Mushroom* son aquellas que disponen de una gran cabeza. Normalmente se encuentran en adultos. Las espinas *Stubby* carecen de cuello y son particularmente prominentes durante el desarrollo postnatal, aunque también se encuentran entre los adultos. Las *Filopodia* son alargadas y normalmente no tienen una cabeza claramente distinguible.

Estudios previos sobre las espinas han conseguido relacionar su morfología con algunas de sus funcionalidades. Por ejemplo se ha sugerido que las espinas finas podrían contribuir al aprendizaje mientras que las más grandes son más estables y podrían estar implicadas en la memoria. Otro objetivo que se cree que pueden cumplir, consiste en incrementar la superficie del área dendrítica con el objetivo de

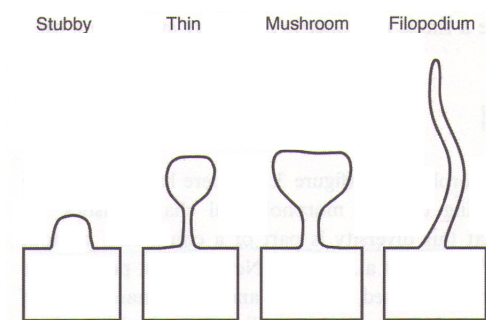


Figura 1.3: Espinas a lo largo de la dendrita. Fuente: Spruston (2008)

1. INTRODUCCIÓN

agrupar un gran número de sinapsis. También cumplen un rol de regulación de las propiedades eléctricas de la neurona.

A la vista del papel que desempeñan las espinas dendríticas en la transmisión sináptica no es de extrañar que un gran número de enfermedades mentales humanas se asocien con alteraciones en su morfología o densidad (Jacobs et al., 1997). Algunos de estos trastornos son la esquizofrenia, en la que la densidad de espinas en las neuronas piramidales neocorticales se sitúa por debajo de la media, otro es el envejecimiento, cuyo estudio se ha centrado principalmente en las neuronas piramidales neocorticales y en las que se ha observado que los sujetos de más de 50 años mostraban un decremento de entre el 9 y 10 % en la longitud total de sus dendritas y una reducción cercana al 50 % en el número de espinas respecto a los individuos de menos de 50 años. También en el atraso mental, que muestra una menor densidad de espinas en el neocortex y el hipocampo y en espinas anormalmente cortas y largas. Los aquí mostrados son algunos ejemplos para ilustrar la importancia de estos componentes en el funcionamiento cerebral ya que la lista de desórdenes asociados con los cambios en las espinas dendríticas es larga y sigue creciendo.



Figura 1.4: Espinas a lo largo de la dendrita.

Fuente: Spruston (2008)

1.2. Objetivos

Con el desarrollo de este proyecto se persigue el estudio morfológico del soma y de las espinas desde el punto de vista de la neuromorfología teórica. El trabajo se basa en un enfoque matemático con el que alcanzar un modelo de representación para ambos componentes neuronales.

Es común en el estado del arte que en el análisis de las características morfológicas de una neurona en tres dimensiones el soma sea ignorado o, en el mejor de los casos, que sea sustituido por una simple esfera. Esto implica que la descripción no coincida con la morfología real del cuerpo celular. Es por ello que se persigue la caracterización de la forma somática mediante un nuevo conjunto de medidas. Para ello se parte de un conjunto de archivos representando el núcleo neuronal junto con sus dendritas primarias. Esto supone que por primera vez se realice una definición matemática del soma con la que determinar objetivamente dónde acaba el soma y

comienza la dendrita. El soma tiene que ser corregido y reparado para posteriormente parametrizar sus propiedades morfológicas. En disposición de las características se procede con la categorización del cuerpo celular en distintas clases. Con ello se discute la existencia de distintos subtipos y se descubren las diferencias entre los somas piramidales de distintas capas del cerebro. Finalmente, a partir del modelo matemático se simulan somas artificiales.

Por su parte, en lo que a las espinas respecta, algunas morfologías de espinas han sido atribuidas a ciertos comportamientos cognitivos. Por ello resulta de interés dictaminar las clases existentes y relacionarlas con las labores de la actividad cerebral. A pesar de la existencia de una categorización ya extendida (Peters y Kaiserman-Abramof, 1970), esta presenta una definición ambigua y subjetiva dependiente de cada individuo y por tanto discutible. Nuestro estudio se sustenta en un conjunto de datos extraídos mediante una técnica novedosa de análisis topológico local para representaciones 3D. Sobre estos datos se trata de alcanzar el conjunto de clases más adecuado en el que agrupar las espinas así como de describir cada grupo mediante reglas que caractericen unívocamente las categorías de las espinas. A partir de los resultados, se discute la existencia de un continuo de espinas y las propiedades que caracterizan a cada subtipo de espina.

1.3. Planificación de la tesis

El desarrollo del proyecto se ha dividido en subobjetivos de acuerdo a las nueve fases descritas en Fayyad et al. (1996) para el análisis de datos hasta convertirlos en conocimiento (ver Figura 1.5).

1. Profundizar en el dominio del problema y dictaminar la meta que se persigue con el proceso de análisis de datos.
2. Generar un conjunto de datos (dataset).
3. Pre-procesar los datos. Esto supone la interpretación del formato de los datos, la definición de las estructuras de datos para su gestión y labores de corrección y eliminación de ruido y errores.
4. Transformar los datos a propiedades más adecuadas desde el punto de vista de la minería de datos y la interpretación de los resultados.

1. INTRODUCCIÓN

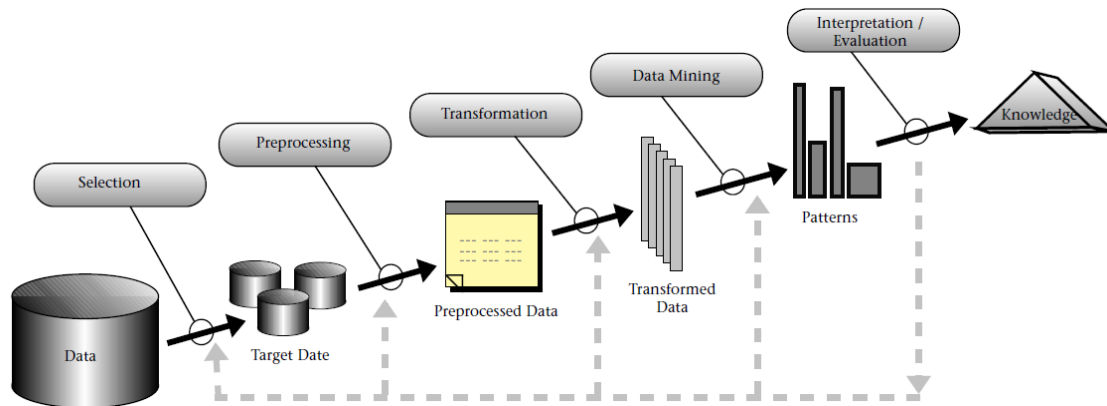


Figura 1.5: Fases del análisis de datos. Fuente: Fayyad et al. (1996)

5. Seleccionar de un método para el análisis de datos adecuado para cumplir los objetivos del estudio.
6. Realizar un análisis exploratorio de los datos y elegir un algoritmo concreto del método escogido en el paso anterior.
7. Detectar de patrones de interés.
8. Interpretar de los patrones mediante representaciones gráficas y análisis estadístico.
9. Incorporar y contrastar el conocimiento adquirido con el existente para incorporarlo a un sistema.

Parte I

Soma

2

Extracción de los datos y pre-procesamiento de la neurona

El concepto de soma es abstracto porque no se dispone de una definición estricta y robusta que especifique exactamente dónde finaliza y comienzan las dendritas. A menudo es tarea de un experto diferenciar entre ambos componentes y cortar la neurona de acuerdo a su criterio. Sin embargo, las medidas morfológicas dependen directamente de los cortes que se realicen y varios expertos segmentando la misma neurona podrían obtener distintos somas. Además, durante el proceso de tinción, la superficie de la neurona es dañada por la inyección agujereándola y deformándola. Así, los somas teñidos no son adecuados para su análisis morfológico porque las medidas en una superficie dañada son incorrectas. En este capítulo se propone un procedimiento para reparar la superficie de un soma tridimensional virtualizado y un método matemático que ofrezca una definición unívoca, justificable y objetiva de qué es un soma.

2.1. Extracción de los datos

Para tal propósito se extrajeron un total de 60 somas junto con sus dendritas primarias repartidos entre el área cingular (25), el área frontal (17) y el área temporal (18). Estas neuronas fueron obtenidas en la autopsia (2-3h post-mortem) y subsecuentemente teñidas con Lucifer Yellow (LY) tras ser inyectadas intracelularmente (Figura 2.1). Las imágenes de las secciones fueron tomadas con un láser confocal de barrido Leica TCS 4D unido a un microscopio fluorescente Leitz DMIRB. Con ello se fotografiaron perfiles de marcaje fluorescente utilizando una longitud de onda de

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

excitación de 491nm. Todo esto permitió obtener un conjunto de imágenes ($\times 20$) que incluyen los árboles apical y basal.

Para cada neurona se capturó la isosuperficie de la malla imponiendo un umbral. Para lograr una superficie sólida que concidiera exactamente con el contorno de la neurona se utilizó el software IMARIS (Bitplane, Zurich, Suiza) con el que se obtuvo una malla triangular, esto es, un conjunto de triángulos que representaban la superficie de la neurona en tres dimensiones. A continuación, cada neurona fue exportada a un fichero VRML. El fichero se compuso de dos elementos básicos, los vértices que definían puntos tridimensionales cartesianos y las caras que denotaron los arcos entre vértices. Cada cara era un triángulo de la malla triangular.

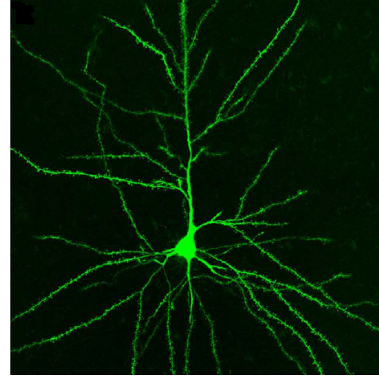


Figura 2.1: Tinción del soma Fuente:(Brito et al., 2013)

Las superficies neuronales almacenadas en el fichero VRML mostraron defectos tales como agujeros y deformaciones introducidos por el proceso de tinción (véase Figura 2.2). El procedimiento de reparación y segmentación se resumió en la Figura 2.3 y se explica a continuación. MeshLab (Cignoni) fue utilizado tanto para el proceso de reparación como para la segmentación.

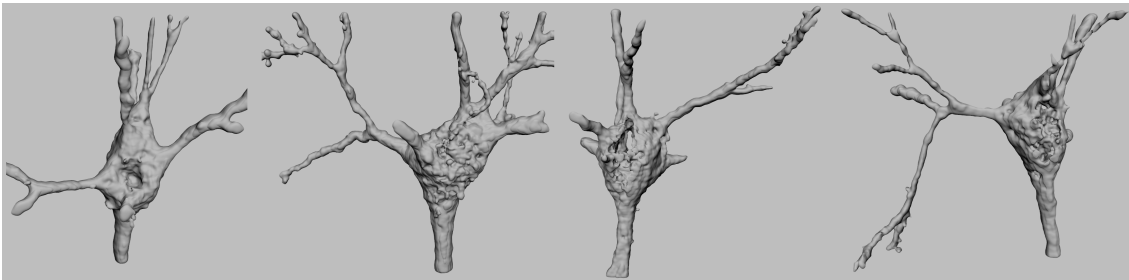


Figura 2.2: Estado inicial de los somas. Cuatro neuronas con algunos defectos en su superficie como agujeros o deformaciones. Estos somas no son adecuados para el análisis morfológico porque las medidas en una superficie dañada son incorrectas.

2.2. Reparación del soma

Las deformaciones en la superficie fueron consideradas como ruido que debía de ser eliminado. El primer paso en el proceso de reparación consistió en distinguir los

vértices en la superficie de la malla de los vértices que formaban los agujeros o se encontraban situados en el interior de la neurona.

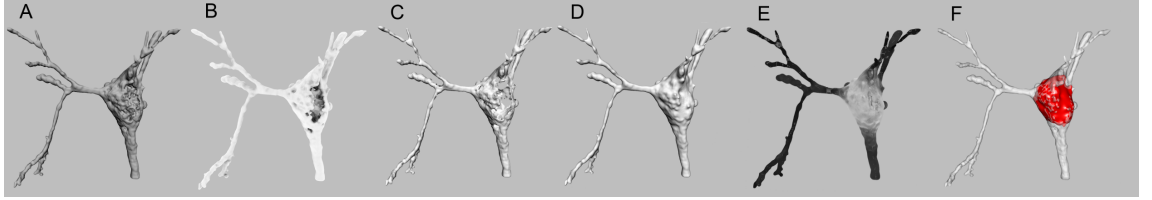


Figura 2.3: Proceso de reparación y segmentación. (A) Estado inicial. (B) La neurona expuesta a la luz ambiental. (C) La neurona una vez que los vértices que formaban los agujeros o se encontraban en el interior de la malla fueron descartados. (D) La neurona tras la reconstrucción de la superficie. (E) Vértices de la malla coloreados de acuerdo a la *Shape Diameter Function*. Segmentación del soma y las dendritas. (F) Resultado final.

Asumiendo que la neurona pudiera ser aislada en un espacio iluminado ficticio, los vértices de la neurona en la superficie de la malla estarían expuestos a la luz mientras que aquellos vértices que formasen un agujero o se encontrasen situados en el interior de la malla quedarían oscurecidos. Así, la información sobre la exposición a la luz podría ser suficiente para distinguir entre los vértices originales en la superficie y los introducidos por la inyección.

Esto motiva el uso de la oclusión ambiental (Favera y Celes, 2012), una técnica que trata de estimar la cantidad de luz que se proyecta sobre un punto de la malla mediante el trazado de rayos. El factor de oclusión ambiental A es una medida de la luz bloqueada por los objetos alrededor del vértice evaluado. Para cada vértice se generó una hemiesfera de radio infinito orientada de acuerdo a su propio vector normal. A continuación, N puntos de la hemiesfera fueron muestreados uniformemente. Después, se trazaron rayos desde el vértice evaluado hacia

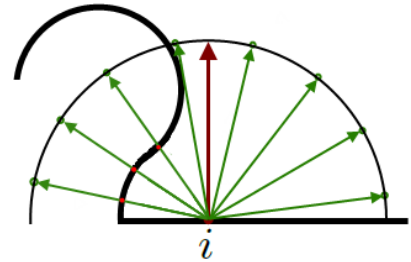


Figura 2.4: Ejemplo de oclusión ambiental en 2D. Hipotética superficie trazada en color negro grueso. En marrón se representa un vértice de la malla y su vector normal. Alrededor del vector normal se genera la hemiesfera en color negro y trazo más fino. En color verde se representan los N puntos muestreados de la hemiesfera y en color rojo las N_c intersecciones de los rayos con la superficie de la malla. Para este caso $A_i = \frac{3}{8}$

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

los puntos muestreados. Contando el número de rayos que intersectaron la superficie de la malla (N_c), obviamente sin considerar el vértice del que surgen los rayos, y comparándolo con el número total de rayos trazados (N), la oclusión ambiental para cada vértice i se calculó como $A_i = \frac{N_c}{N}$ (Figura 2.4). Como resultado, se obtuvo un valor escalar A para cada vértice de la malla. El dominio de A_i es $[0, 1]$, donde 0 denota que ningún rayo intersectó la superficie de la malla y 1 implica que todos los rayos trazados intersectaron la malla y que, en consecuencia, el vértice se encontraba en el interior de la malla. Así, se obtuvo la exposición de la neurona a la luz ambiental de acuerdo a los valores de A (Figuras 2.3B 2.5).

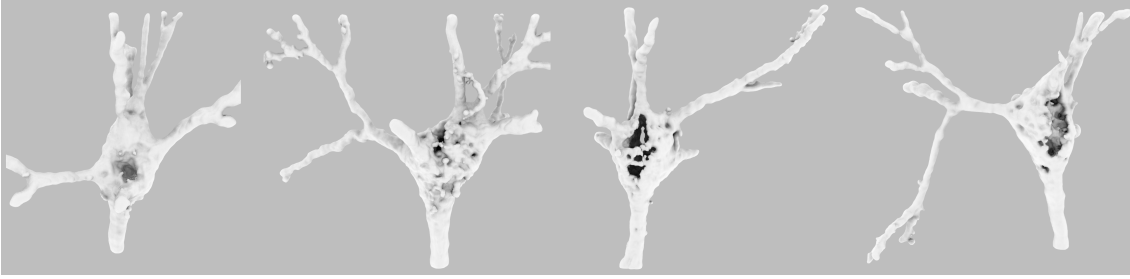


Figura 2.5: Oclusión ambiental aplicada a la neuronas. Exposición de la neurona a la luz ambiental de acuerdo al valor escalar A denotado por la oclusión ambiental para cada vértice. Aquellos puntos cuyo A_i fue cercano a 0 fueron coloreados de blanco mientras que aquellos puntos cuyo A_i fue cercano a 1 fueron coloreados de negro. Se muestra que los puntos de la superficie se encontraron expuestos a la luz (blanco) mientras que los vértices en el interior fueron oscuros (negro).

Se encontró otro problema, debido a que algunos vértices fueron introducidos durante el proceso de tinción estos tuvieron que ser eliminados. Para ello, un enfoque sencillo podría consistir en imponer un umbral arbitrario para los valores escalares A tal que los vértices con un A superior al umbral fueran descartados. Sin embargo, la estimación de un umbral debería de ser preferiblemente automática, justificable y objetiva.

En esta situación se consideró que un algoritmo de clustering, cuya meta es agrupar instancias de datos similares en el mismo grupo, concordaba con las especificaciones del problema. Se aplicó un clustering de mixtura de Gaussianas (McLachlan y Basford, 1988) para clasificar los vértices en dos grupos: por un lado, los vértices de la superficie de la neurona y por otro lado tanto los vértices en el interior de la neurona como los que formaban agujeros. El clustering de mixtura de Gaussianas devolvió para cada vértice la probabilidad de pertenencia a cada cluster. El umbral

fue definido como el límite de decisión entre los dos clusters, esto es, el valor de A en el que ambos grupos fueron equiprobables. Los vértices cuyo valor A fue más alto que el umbral fueron eliminados así como las caras asociadas con ellos. En consecuencia la malla fue agujerada como se puede apreciar en las Figuras 2.3C y 2.6 y por lo tanto tuvo que ser de nuevo reconstruida.

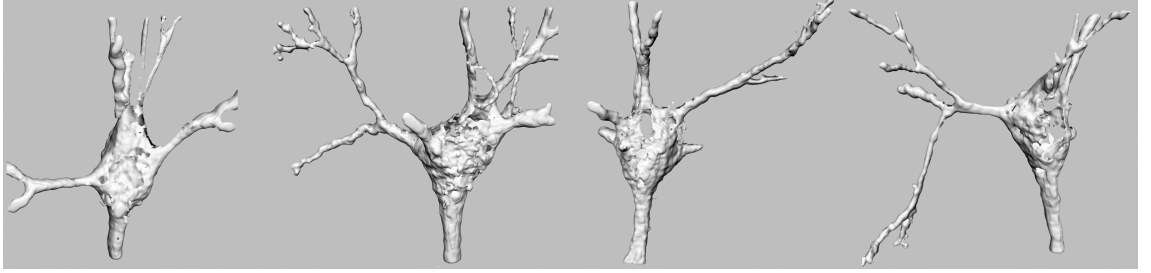


Figura 2.6: Neuronas agujereadas después de eliminar los vértices por encima del umbral.

2.2.1. Reconstrucción de la malla

Una manera sencilla de definir la superficie de un objeto es mediante una función indicadora que asigne 1 a los puntos en el espacio interior de la malla y 0 a los puntos en el espacio fuera de la malla (Figura 2.7A). De esta forma, la frontera entre 1s y 0s determina la superficie del objeto. Asumiendo que el espacio es continuo y que ha sido evaluado mediante la función indicadora, si se calcula el gradiente de ese espacio, como resultado se mostraría que solamente los puntos en la frontera serían distintos de 0. Aquellos puntos del gradiente con un valor diferente de 0 definirían una curva cerrada que denotaría una isosuperficie (Figura 2.7B). La isosuperficie se correspondería con la superficie del objeto.

Sin embargo, no se dispone de tal función indicadora y solamente se tiene acceso a los puntos de la malla y a las normales interiores de las caras, información provista por Imaris y MeshLab (Figura 2.7C). Por lo tanto, la función indicadora tuvo que ser calculada a partir de los datos de la malla. En Kazhdan et al. (2006) se propone y se explota la relación entre el gradiente de la función indicadora y la integral del campo producido por las normales de la superficie. Como la integral no puede ser evaluada porque se dispone de un conjunto discreto de vértices y caras, se discretizó el espacio de acuerdo a particiones o parcelas. El valor de la integral para una parcela fue aproximado de acuerdo a los valores del campo vectorial producido por las normales

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

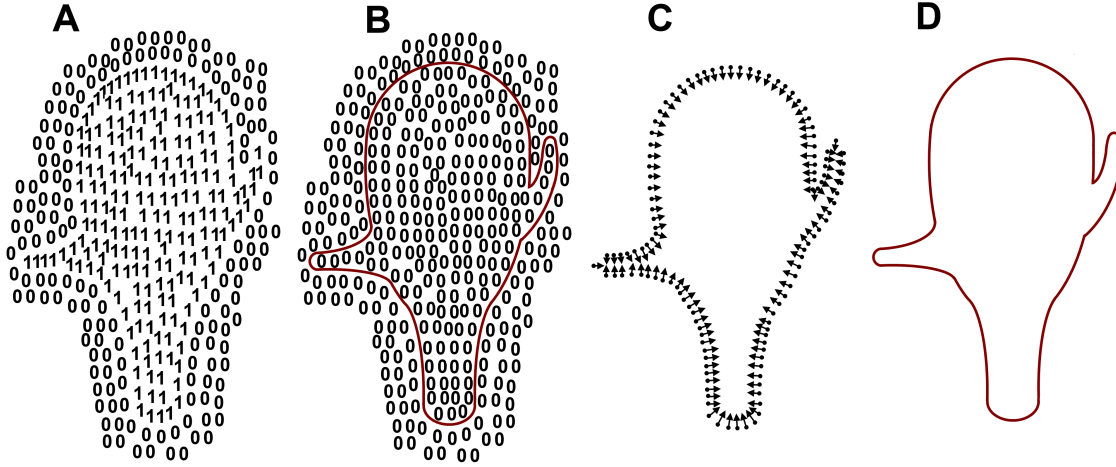


Figura 2.7: Reconstrucción para mallas. (A) La función indicadora. El espacio en el interior de la malla se denota mediante un 1 y el ubicado en el exterior con un 0. (B) Gradiente de la función indicadora. Dado que la función indicadora es constante en el exterior (0s) y en el interior (1s) de la malla, el gradiente de esos puntos es 0. Solo los puntos en la frontera son distintos de 0 lo que define una isosuperficie (rojo). (C) Campo vectorial definido por los puntos orientados. (D) Superficie de la malla. Adaptado de Kazhdan et al. (2006)

de los vértices dentro del área de la parcela, escalando el valor respecto al volumen de la parcela. A partir de los valores de las parcelas se obtuvo una isosuperficie (ver Figura 2.7D). Si un usuario incrementa el número de parcelas, la isosuperficie se aproxima con más detalle. Sin embargo, un número muy grande de parcelas podría introducir ruido, especialmente cuando la cantidad de vértices y caras es limitada. En el caso contrario, cuando hay un número pequeño de parcelas, la superficie de la malla se suaviza y se pierden algunos detalles.

Un ejemplo del resultado tras la reconstrucción de las neuronas se muestra en las Figuras 2.3D y 2.8.

2.3. Segmentación automática de la neurona

La segmentación de la neurona puede ser entendida como un problema de clustering donde cada vértice pertenece a un grupo, bien a los somas o bien a las dendritas. Sin embargo, no se trata de un problema trivial debido a que el soma es un concepto abstracto y no existe una definición estricta y robusta que establezca con exactitud dónde acaba y comienzan las dendritas. De hecho, normalmente es tarea de un

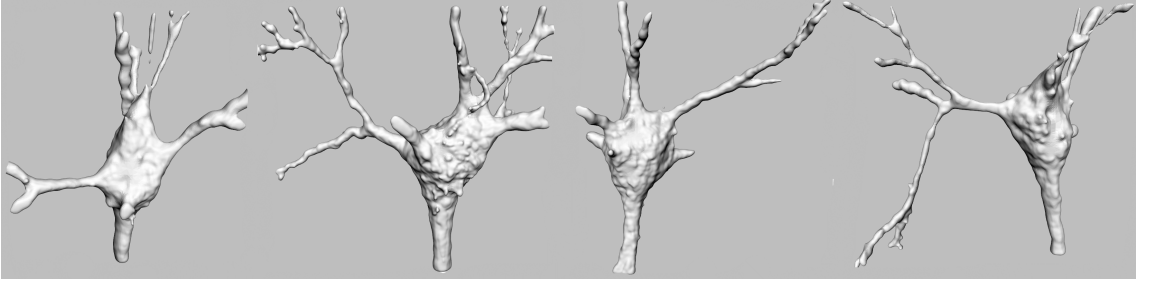


Figura 2.8: Reconstrucción de las neuronas. Tras la aplicación del proceso de reconstrucción se han eliminado los agujeros producidos tras la oclusión ambiental y además se ha suavizado la superficie del soma ligeramente.

experto diferenciar entre ambos componentes y cortar la neurona de acuerdo a su criterio. Sin embargo, las medidas morfológicas dependen directamente de los cortes y varios expertos segmentando la misma neurona podrían obtener diferentes somas. Esto motiva un modelo matemático para obtener un resultado unívoco, justificable y objetivo.

En Shapira et al. (2008) se presenta una función escalar, *Shape Diameter Function* (SDF), que trata de explotar las diferencias del volumen en la vecindad de los vértices de una malla. Esta propiedad es adecuada para la segmentación dado que las dendritas son más finas que los somas y en consecuencia el volumen en la vecindad de los vértices del soma es más grande.

Para cada vértice se calcula la SDF. Básicamente, dado un vértice de la malla y un cono centrado en la dirección de su normal hacia el interior de la malla, se muestrean va-

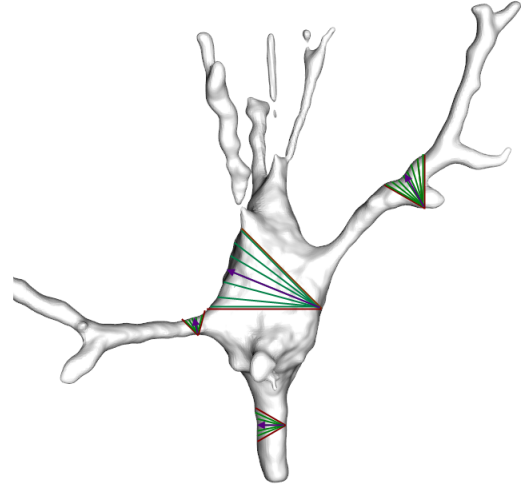


Figura 2.9: Ejemplo de *Shape Diameter Function*. En la imagen se muestra un ejemplo de cómo se calculan los valores escalares para cada vértice de la malla tridimensional. El cono (marrón) se centra en la normal hacia el interior de la malla (flecha morada). Varios rayos (verde) se muestrean en el interior del cono. Los rayos muestreados en el interior del soma son más largos que los rayos muestreados en el interior de las dendritas por lo que el volumen en la vecindad de los vértices del soma es más alto.

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

rios rayos en el interior del cono tales que la suma de la longitud de los rayos desde el vértice hasta su intersección con la superficie de la malla en el lado opuesto aproxime el volumen en su vecindad (Figura 2.9). A partir de la salida de la SDF se obtuvieron las neuronas coloreadas como se muestra en la Figura 2.3E y 2.10.

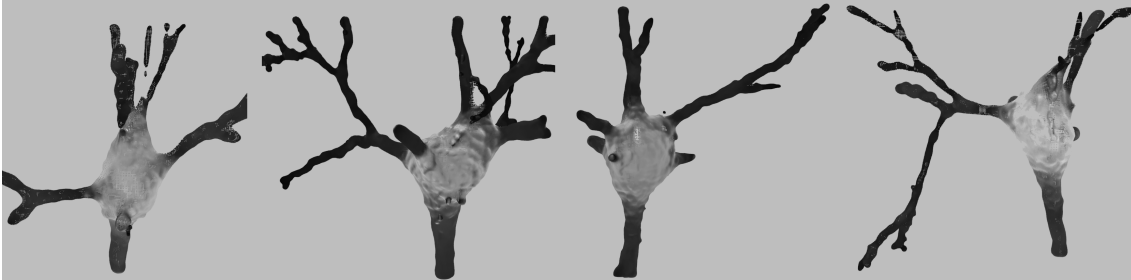


Figura 2.10: SDF aplicado a neuronas. Los vértices de la malla son coloreados de acuerdo al valor de la función escalar SDF. Cuanto más oscuro es el vértice, más pequeño es el volumen de la vecindad. En consecuencia, los vértices de los somas son grises y los vértices de las dendritas son negros.

Como con la oclusión ambiental, algunos vértices tuvieron que ser eliminados. En este caso, los vértices del soma se mantuvieron mientras que los vértices de las dendritas fueron descartados. Así, se impuso un umbral de acuerdo a la salida de la SDF. De nuevo se aplicó un clustering probabilista basado en mixturas finitas de gaussianas con el fin de alcanzar un modelo matemático para la agrupación de vértices. Aparentemente, la salida de la SDF presentó una distribución unidimensional compuesta por una mixtura de dos Gaussianas (Figura 2.11), el soma y las dendritas. Sin embargo, en varios casos la dendrita apical fue más gruesa que las dendritas basales y el algoritmo de clustering consideró que la dendrita apical era parte del soma. Cuando se aplicó el clustering en busca de tres clases, en aquellos casos en los que la dendrita apical y las dendritas basales fueron bastante similares, el soma fue cortado por la mitad. El algoritmo de clustering tuvo problemas para agrupar los vértices de la dendrita apical debido a que la Gaussiana que representaba a la dendrita apical se ocultaba con la Gaussiana del soma y la Gaussiana de las dendritas basales.

Con el objetivo de resolver el problema se definió un proceso de dos etapas. Para comenzar, se aplicó un clustering con dos clases de acuerdo a la distribución de la SDF (Figura 2.11) para distinguir entre los vértices de las dendritas basales y los vértices del soma y la dendrita apical. A continuación, los vértices de las dendritas

2.3 Segmentación automática de la neurona

basales fueron eliminados (Figura 2.12). Para la segunda etapa se aplicó de nuevo el clustering pero esta vez sobre la salida del primer clustering, esto es, sobre los vértices del soma y de la dendrita apical (Figura 2.13). Como resultado del proceso los vértices de la dendrita apical fueron eliminados y el soma fue identificado. Cuando los vértices de la dendrita apical y los vértices de las dendritas basales eran bastante parecidos el segundo clustering aumentó la precisión de los cortes.

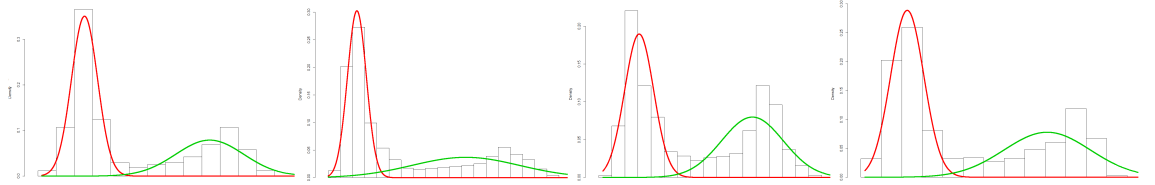


Figura 2.11: Histograma y primer clustering. En las imágenes se representan la distribución del volumen del soma y la dendrita apical (rojo) y las dendritas basales (verde). Aparentemente hay dos Gaussianas. Sin embargo, en la segunda representación de la Figura 2.13 hay dos Gaussianas lo que demuestra que la dendrita apical se encontraba oculta.

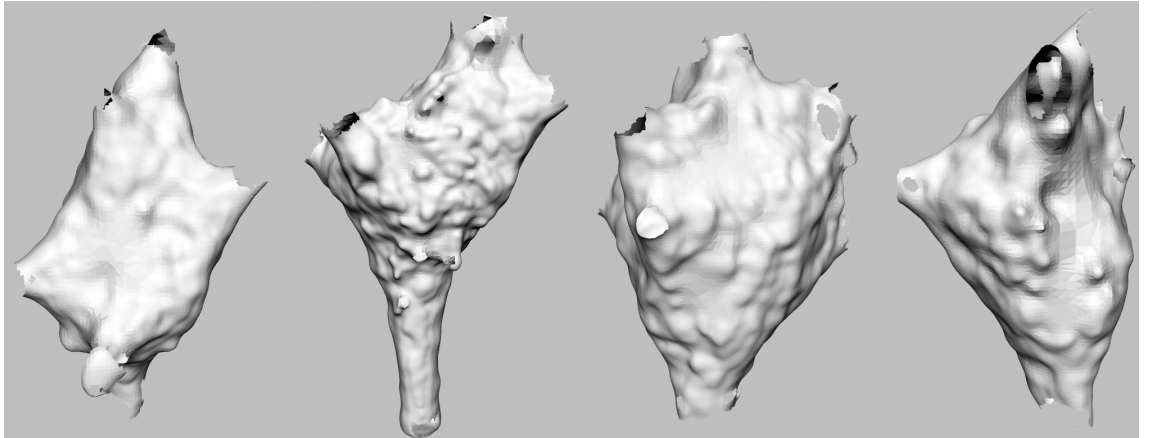


Figura 2.12: Neuronas tras el primer clustering. Como se puede observar, el clustering con dos grupos (Figura 2.11) en algunos casos no eliminó la dendrita apical. Esto motivó un segundo clustering.

Tras este proceso la malla tridimensional fue agujereada y reconstruida. Un ejemplo del final del proceso se muestra en las Figuras 2.3F y 2.14.

Al final del procedimiento de reparación y segmentación se pudieron recuperar un total de 40 somas ya que los 20 restantes se encontraban demasiado deteriorados

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

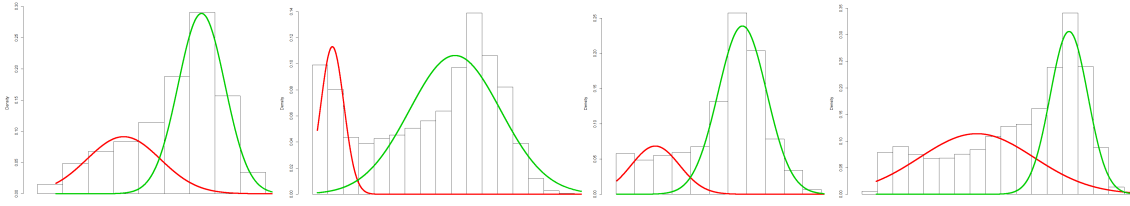


Figura 2.13: Histograma y segundo clustering. Se muestra la distribución del volumen del soma (rojo) y la dendrita apical (verde). En algunos casos este clustering sirve para eliminar vértices de la dendrita apical como en la segunda imagen y en otros para mejorar el resultado del primer corte como en la cuarta imagen.

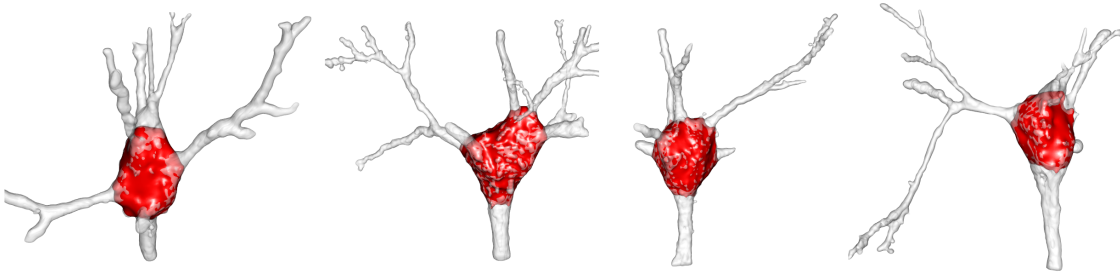


Figura 2.14: Resultado final. Soma reparado y extraído (rojo) sobre la neurona original mostrada en la Figura 2.2.

como para que el proceso de reparación fuera efectivo en ellos.

2.3.1. Bondad del método de segmentación del soma

Con el fin de validar la bondad del método de segmentación automática de somas se tomaron como referencia varios somas segmentados por expertos. De esta forma se compararon los cortes de los neurocientíficos con los realizados por la herramienta. En la Tabla 2.1 y en la Figura 2.15 se presentan los resultados. Además, también se compararon los cortes de un mismo experto sobre un mismo soma en días distintos. Los resultados se muestran en la Tabla 2.2 y la Tabla 2.3. Como se observa, la herramienta obtuvo definiciones de soma semejantes a la de los expertos pero con la aplicación del software se dispuso de algunas ventajas como son la objetividad y la rapidez en el procedimiento.

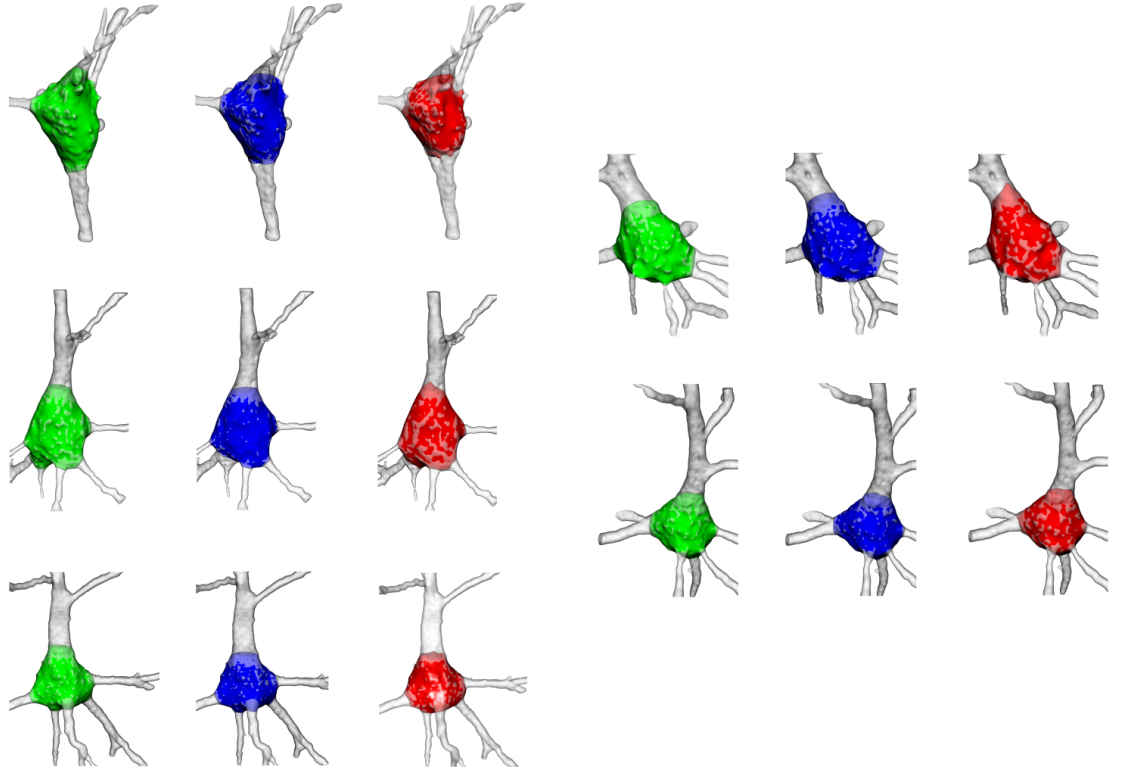


Figura 2.15: Bondad del método de segmentación del soma. En color verde y azul se presentan los somas segmentados por los expertos 1 y 2 mientras que en color rojo se presentan los somas segmentados por la herramienta.

	Exp.1 Vs Exp. 2	Exp. 1 Vs Algoritmo	Exp. 2 Vs Algoritmo
cingidm16portalsecc1somacel1	0.28	0.75	0.82
cingidm16portalsecc1somacel2	0.39	0.56	0.50
cingidm16portalsecc1somacel3	0.35	0.81	0.72
cingidm16portalsecc1somacel5	0.43	0.33	0.20
cingidm16portalsecc1somacel6	0.57	1.34	0.69
tempidm16portalsec1capa3cel10	0.97	0.50	0.72
tempidm16portalsec1capa3cel9	0.49	0.50	0.47
tempidm16porta3sec1capa3cel22	0.40	0.55	0.31
tempidm16porta3sec1capa3cel25	0.31	0.21	0.29

Tabla 2.1: Diferencias en la segmentación del soma de expertos y del algoritmo. La tabla representa la raíz de la desviación cuadrática media (RMSE) de la mínima distancia entre el soma segmentado por un experto y el soma segmentado por el otro experto o el algoritmo. Dado que esta medida es asimétrica se calculó el recíproco y se seleccionó el peor de los dos resultados. En negrita se denotan aquellos casos en los que el RMSE fue menor. El algoritmo se aproximó más al corte de un experto (5 veces) que entre los dos expertos (4 veces).

2. EXTRACCIÓN DE LOS DATOS Y PRE-PROCESAMIENTO DE LA NEURONA

	Día 1 Vs Día 2	Día 1 Vs Día 3	Día 2 Vs Día 3
cingidm16portalsecc1somacel1	0.51	0.27	0.65
cingidm16portalsecc1somacel3	0.26	0.21	0.39
cingidm16portalsecc1somacel5	0.27	0.12	0.20
cingidm16portalsecc1somacel6	0.99	0.70	0.35
tempidm16portalsec1capa3cel9	0.79	0.51	0.25
tempidm16porta3sec1capa3cel22	0.35	0.44	0.48

Tabla 2.2: Diferencias en la segmentación del soma del experto 1 sobre el mismo soma distintos días. Esta tabla es similar a la Tabla 2.1 pero en este caso compara las diferencias del RMSE de un mismo experto cortando los mismos somas 3 días distintos. En negrita se muestran las mayores diferencias. Se puede observar como varía la segmentación a pesar de que el corte es realizado por el mismo experto sobre los mismos somas lo que motiva el uso de la herramienta.

	Día 1 Vs Día 2	Día 1 Vs Día 3	Día 2 Vs Día 3
cingidm16portalsecc1somacel1	0.19	0.51	0.50
cingidm16portalsecc1somacel3	0.25	0.40	0.19
cingidm16portalsecc1somacel5	0.20	0.28	0.13
cingidm16portalsecc1somacel6	0.62	0.56	0.37
tempidm16portalsec1capa3cel9	0.34	0.68	1.03
tempidm16porta3sec1capa3cel22	0.73	0.40	0.32

Tabla 2.3: Diferencias en la segmentación del soma del experto 2 sobre el mismo soma distintos días. Tabla similar a la Tabla 2.2 pero en este caso con los resultados del experto 2. En negrita se denotan las mayores diferencias. De nuevo se demuestra la variación en los cortes de un mismo experto sobre los mismos somas por lo que se justifica el uso de la herramienta presentada en este capítulo.

3

Caracterización del soma

Al finalizar el pre-procesamiento expuesto en el capítulo anterior se obtuvo un conjunto de 40 somas representados mediante mallas triangulares. Para alcanzar una comprensión más profunda sobre su morfología los vértices y caras tuvieron que ser transformados a un conjunto de propiedades. Como es habitual, para determinar el conjunto de características sobre las que fundamentar el análisis se consultó el estado del arte. Las propiedades comúnmente utilizadas para describir al soma resultan de índole genérica y por lo tanto no especifican una morfología concreta sino que varios somas coinciden con una misma descripción. En consecuencia, no son adecuadas para un proceso posterior de simulación. Para caracterizar adecuadamente el soma se propone un método que recoge propiedades tanto locales como globales de la morfología somática. En este capítulo se dictamina el proceso de transformación de los datos y se postula el conjunto de descriptores sobre el que se desarrolla el estudio en fases posteriores.

3.1. Caracterización basada en el estado del arte

El pre-procesamiento de las neuronas dio como resultado un conjunto de 40 somas representado mediante mallas triangulares. Puesto que el estudio de la morfología somática mediante vértices y caras dificulta la interpretación de los resultados, los datos tuvieron que ser transformados a propiedades más fácilmente comprensibles. Con este fin, se ahondó en las propiedades de uso más común en el estado del arte para la caracterización de somas.

Los descriptores recogidos en la literatura fueron divididos en dos grupos. En el primero de ellos se analizaron medidas comunes mientras que en el segundo se

3. CARACTERIZACIÓN DEL SOMA

estudió la semejanza entre el soma y los sólidos platónicos. Con estos subconjuntos se pretendió por un lado discutir la existencia de distintos tipos de cuerpos somáticos y por otro lado evaluar su piramicidad. A continuación se exponen las características.

3.1.1. Descriptores globales

1. **Volumen** (Harland, 2013; Masseroli et al., 1993; Meijering, 2010): Para su cálculo se utilizó el método propuesto en Cha y Tsuhan (2001). Se genera un tetraedro para cada cara uniendo los tres vértices del triángulo con el origen de coordenadas. La suma de los volúmenes de los tetraedros con signo (positivo o negativo) da como resultado el volumen total de la malla. El signo se determina a partir de la normal de la cara. El cálculo del volumen para una cara/tetraedro i es:

$$V_i = \frac{1}{6}(-x_{i3}y_{i2}z_{i1} + x_{i2}y_{i3}z_{i1} + x_{i3}y_{i1}z_{i2} - x_{i1}y_{i3}z_{i2} - x_{i2}y_{i1}z_{i3} + x_{i1}y_{i2}z_{i3}),$$

donde x , y y z son las coordenadas de un vértice y su subíndice determina el vértice. El volumen total de la malla es la suma de todos los volúmenes de todos los tetraedros:

$$V_{total} = \sum_i V_i.$$

2. **Superficie del área** (Alavi et al., 2009; Harland, 2013; Masseroli et al., 1993; Meijering, 2010): Este valor fue calculado como la suma de todas las áreas de los triángulos que conforman la superficie. Dado que el cálculo de la longitud de los tres lados del triángulo es trivial dados los vértices del triángulo, se aplicó la fórmula de Herón (de Alejandría, 60):

$$A_i = \sqrt{s_i(s_i - x_i)(s_i - y_i)(s_i - z_i)},$$

donde x, y y z son los vértices de la cara i y s_i , el semiperímetro del triángulo, se obtiene a partir de:

$$s_i = \frac{1}{2}(x_i + y_i + z_i).$$

El área total de la malla se calcula mediante:

$$A_{total} = \sum_i A_i.$$

3. **Diámetro máximo o de Feret** (Harland, 2013; Masseroli et al., 1993; Meijering, 2010): La definición de este descriptor insta a la obtención de la máxima distancia entre los vértices que componen el soma. Su cálculo se basa en encontrar los puntos p_i y p_j de la malla que cumplen $\arg \max_{ij} |\vec{d}(p_i, p_j)|$ siendo \vec{d} la distancia euclídea entre puntos. Dado que calcular la distancia entre todos los n puntos tiene una complejidad $\mathcal{O}(n^2)$, para acelerar el proceso se calculó primero la envoltura convexa de los puntos y, sobre los puntos que conformaban la envoltura convexa, se seleccionó el par que cumpliera que maximizase la distancia.
4. **Diámetro menor** (Harland, 2013; Masseroli et al., 1993; Meijering, 2010): Para su cálculo la malla se situó en una caja de tamaño mínimo o “bounding box”. A partir de las dimensiones de la caja se obtuvieron las extensiones de la malla en X, Y y Z.
5. **Cociente entre ejes** (Masseroli et al., 1993; Meijering, 2010): Es la proporción entre el eje mayor y el eje menor, esto es, $\frac{D_{max}}{D_{min}}$. En este caso el ratio la longitud mayor y la longitud menor de la “bounding box”.
6. **\log_{10} del cociente de los ejes** (Dumitriu et al., 2007; Tsiola et al., 2003).
7. **Número de vértices** (Alavi et al., 2009). Número total de vértices de la malla tridimensional.
8. **$\frac{4\pi \times \text{volumen}}{\text{Area}^2}$** : Se trata de una medida propia de Kozlowski y Weimer (2012), que ha sido extrapolada aquí de 2D a 3D.
9. **Distancia media desde el centro de masas a los vértices** (Bélisle et al., 2012): Para su cálculo en primer lugar se obtiene el centro de masas de la malla mediante $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N V_i$, donde N es el número de vértices y V_i es el vértice i . A continuación se computa la distancia media mediante $\hat{d} = \frac{1}{N} \sum_{i=1}^N |\vec{d}(\hat{\mu}, V_i)|$.
10. **Desviación estándar desde el centro de masas a los vértices** (Zhang et al., 2012). Este valor se calcula a partir de la desviación estándar entre el centro de masas de la malla, valor que se obtuvo para el punto anterior, y cada uno de los vértices de la malla. Por lo tanto $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_i - \hat{\mu})^2}$.
11. **Solidez** (Benavides-Piccione et al., 2006): Área de la superficie de la malla dividida por el área de la envoltura convexa.

3. CARACTERIZACIÓN DEL SOMA

12. **Elongación** (Masseroli et al., 1993; Meijering, 2010): Es la inversa de la esfericidad cuyo cómputo se expone posteriormente.
13. **Elípticidad** (Alavi et al., 2009; Masseroli et al., 1993; Meijering, 2010): Se trata de una medida de compresión de un círculo o una esfera basada en los diámetros. Su cálculo es $\frac{\pi \times \frac{D_{max}}{2} \times \frac{D_{min}}{2}}{area}$.

3.1.2. Descriptores basados en similaridad morfológica

Se definen un conjunto de descriptores que comparan la forma de un objeto con los sólidos platónicos y la esfera.

1. **Esfericidad** (Alavi et al., 2009; Masseroli et al., 1993; Meijering, 2010): El cálculo de la esfericidad de un cuerpo tridimensional se basa en la comparación entre la superficie de un objeto y la superficie de la esfera del mismo volumen. A partir del volumen de la malla se calcula el radio de la esfera r mediante $r = \sqrt[3]{\frac{3V}{4\pi}}$. Dado el radio de la esfera el área se calcula mediante $A_s = 4\pi r^2$. Finalmente, dado el área de la malla (A_m) el cálculo de la esfericidad es $Esfericidad = \frac{A_m}{A_s}$.
2. **Piramicidad** (Masseroli et al., 1993): Se siguió el mismo proceso desarrollado para la esfericidad pero basado en un tetraedro. Se calculó la relación entre el volumen y las aristas del tetraedro de manera que $V = \frac{\sqrt{2} \cdot a^3}{12}$. La longitud de la arista del tetraedro se obtiene como $a = \sqrt[3]{\frac{12 \cdot V}{\sqrt{2}}}$. Si se obtiene la apotema t de la pirámide (altura del triángulo de una de las caras) a partir del teorema de Pitágoras $a^2 = t^2 + (\frac{a}{2})^2$ el área del tetraedro se obtiene mediante $A_p = 4 \cdot \frac{a \cdot t}{2}$ de manera que $Piramicidad = \frac{A_m}{A_p}$.
3. **Cubicidad**: El cálculo de esta variable es similar al de las anteriores. En este caso $A_c = 6 \cdot \sqrt[3]{V^2}$ con lo que $Cubicidad = \frac{A_m}{A_c}$.
4. **Octaedricidad**: Para comparar este sólido platónico con el soma se calculó la longitud de la arista a partir de $a = \sqrt[3]{\frac{3V}{\sqrt{2}}}$. A partir de a , el cómputo de la superficie del octaedro con volumen V es $A_o = 2\sqrt{3} \cdot a^2$. La octaedricidad se obtiene a partir de $Octaedricidad = \frac{A_m}{A_o}$.
5. **Dodecaedricidad**: La longitud de la arista de un dodecaedro en relación con su volumen puede ser obtenida a partir de $a = \sqrt[3]{\frac{4V}{(15+7\sqrt{5}) \cdot a^3}}$. Una vez a es

conocido, el cálculo del área se obtiene a partir de $A_d = 3\sqrt{25 + 10\sqrt{5}} \cdot a^2$. Finalmente $Dodecaedricidad = \frac{A_m}{A_d}$.

6. **Icosaedricidad:** Para este cuerpo poliédrico el cálculo de la longitud de la arista en función a un volumen V se obtiene a partir de $a = \sqrt[3]{\frac{12V}{5(3+\sqrt{5})}}$. Dado el valor de a , el cálculo de la superficie del icosaedro es $A_i = 5\sqrt{3} \cdot a^2$, por lo que $Icosaedricidad = \frac{A_m}{A_i}$.

Al investigar el estado del arte se observó que en la literatura, cuando se realiza un análisis de las características morfológicas de una neurona en tres dimensiones, es común que el soma sea ignorado o, en el mejor de los casos, que sea sustituido por una simple esfera (Quan et al., 2013; Stelescu et al., 2012). Esto puede deberse a la aplicación de la técnica histológica de Golgi-Cox que produce defectos en la superficie somática, defectos que fueron reparados en el pre-procesamiento. Puesto que la información recabada al revisar el estado del arte especifica descriptores con los que no es posible representar una morfología única ya que varios somas coinciden con una misma descripción, no son adecuadas para un proceso posterior de simulación. Debido a estos factores, la información extraída del estado del arte se utilizó como componente de apoyo para descubrir diferencias entre clases de somas y somas en distintas capas, es decir, se aplicaron como variables emergentes. Para caracterizar adecuadamente el soma se desarrolló un método que recoge propiedades tanto locales como globales de la morfología somática.

3.2. Caracterización basada en Reeb graph

Debido al aumento en el número de imágenes tridimensionales existe un interés creciente en desarrollar métodos precisos y eficientes para la búsqueda en bases de datos de imágenes 3D. Con este fin se han desarrollado multitud de técnicas para la extracción de características de figuras tridimensionales. Uno de estos métodos es *Reeb graph* mediante el cual se define una estructura topológica que genera el esqueleto de un objeto en un número de dimensiones arbitraria (Hilaga et al., 2001). Esta técnica construye un grafo de un modelo geométrico 3D para describir la topología del objeto.

La idea básica tras este enfoque es particionar la malla triangular en regiones de acuerdo a una función continua $\mu : C \rightarrow \mathbb{R}$, donde $C \in \mathbb{R}^3$ son los límites del soma. Existen diversas funciones μ aplicadas sobre la malla 3D como son la altura

3. CARACTERIZACIÓN DEL SOMA

relativa entre la posición del punto y la altura total del objeto, la distancia media al centroide del objeto o la distancia geodésica, siendo esta última la de uso más extendido. Estas funciones se calculan para cada uno de los puntos de la malla tridimensional con lo que a cada punto se le asigna un valor. A partir de los valores de la función se obtiene el dominio $\mu(\cdot)$ que se divide en K intervalos r_i de manera que $r_0 = [0, \frac{1}{K}]$, $r_1 = [\frac{1}{K}, \frac{2}{K}]$, \dots , $r_{K-1} = [\frac{K-1}{K}, 1]$. Al dividir la malla tridimensional en intervalos, las fronteras entre cada una de las secciones definidas por los intervalos presentan formas muy diversas y complejas dificultando la interpretación de los resultados. De hecho, en áreas donde la concentración de vértices es reducida, situación que ocurre en algunas localizaciones donde la malla tridimensional fue reconstruida, el problema se acrecenta.

3.2.1. Ray-tracer graph

Los inconvenientes presentados, junto con la morfología presentada por el soma, motivaron el desarrollo de una nueva técnica para la extracción de características a partir de una malla tridimensional. Esta técnica puede ser interpretada como una adaptación de *Reeb Graph* dado que la topología se describe a partir del esqueleto del objeto representado por un grafo que surge de los K intervalos.

Para presentar la solución, se considera la situación ideal en la que el cuerpo es esférico. Si se cumple esta condición, el centro de masas de los puntos que definen la figura determina el centro de un sistema de coordenadas esféricas en el que cada punto puede ser representado a partir de los ángulos de elevación (colatitud) y latitud (azimutal). De acuerdo a este sistema, si se fija la colatitud y se recorren todos los valores del ángulo azimutal $[0, 2\pi]$ se describe un círculo que parte la esfera en dos secciones. Ahora, si la colatitud se fija a otro valor y se repite el proceso con el ángulo azimutal se consigue una esfera dividida en tres partes. El proceso puede repetirse tantas veces como se desee. El ángulo de elevación es, en realidad, una nueva medida para la función μ . Puesto que la morfología del soma es habitualmente considerada una esfera, el método que se ha descrito puede utilizarse para aproximar K secciones dando $K + 1$ valores al ángulo de elevación. La ventaja de este método sobre el tradicional consiste en que, al trazar rayos, las secciones no se delimitan por los vértices de la malla sino que lo hacen las intersecciones de los rayos con las caras. Al trabajar sobre las caras se da una situación ideal de continuidad que ofrece límites entre secciones suavizadas con lo que se superan los inconvenientes presentados por el resto de funciones μ .

Para la aplicación de este nuevo método todos los somas fueron orientados de igual forma haciendo coincidir el primer componente principal, obtenido a partir del análisis de componentes principales (PCA) de los puntos del soma, con una colatitud de 90° . A continuación, puesto que la morfología del soma no tiene por qué coincidir con una esfera perfecta, se siguió un procedimiento basado en el cálculo de la esfera de tamaño mínimo que contuviera al soma. El radio de la esfera mínima quedó determinado por $r = \frac{d_f}{2}$ donde r es el radio de la esfera y d_f el diámetro máximo o de Feret de la malla triangular. Con los somas orientados y contenidos en esferas comenzó el proceso de trazado de rayos. Los puntos de intersección de los rayos emitidos desde el centroide del soma hacia los puntos señalados por los ángulos colatitud y azimuth con la malla triangular determinaron los límites entre secciones. Variando el número de valores para la colatitud y el azimuth se pueden lograr aproximaciones con más o menos detalle de la malla 3D. En la Figura 3.1 se muestra el resultado.

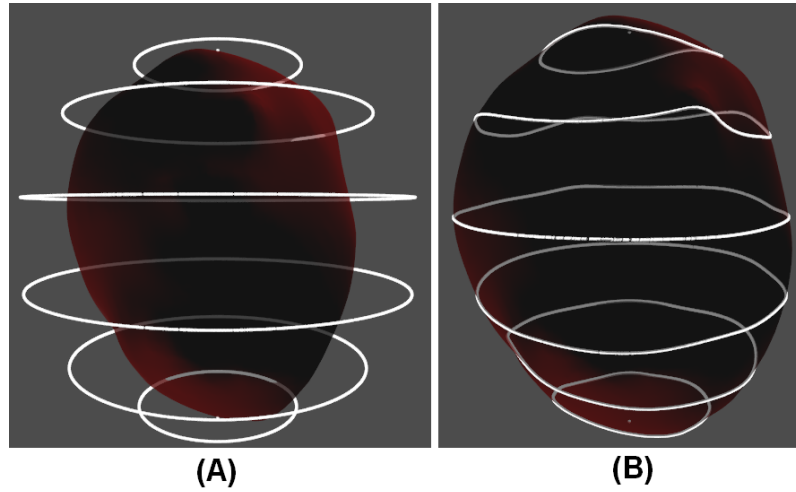


Figura 3.1: Ray-tracer graph. (A) Se genera una esfera de radio $\frac{d_f}{2}$ alrededor del soma. Se determinan K particiones a partir de $K - 1$ circunferencias definidas por $K + 1$ valores para el ángulo colatitud. (B) Se trazan rayos desde el centroide del soma hacia los puntos de cada circunferencia lo que da lugar a intersecciones entre el rayo y la malla que definen el límite entre cada sección.

La aplicación de este método da lugar a límites entre secciones o curvas de diversos perfiles. Esta variedad de formas permitió aproximar con detalle la morfología local del soma. Con el propósito de modelar la morfología del soma, se aproximaron las curvas de manera que con un conjunto reducido de descriptores la diferencia en-

3. CARACTERIZACIÓN DEL SOMA

tre la curva real y la aproximada fuera la mínima. El proceso se dividió en dos fases. En la inicial se estudiaron las curvas desde el punto de vista perpendicular al plano XY . De esta forma la representación bidimensional de las curvas fue aproximada mediante elipses. A continuación se modeló la tercera dimensión mediante las dos primeras.

El primer paso del proceso de ajuste de las curvas se fundamentó en la orientación del soma. El eje principal del soma que se había obtenido mediante PCA se situó sobre el eje cartesiano de cotas de manera que las curvas quedasen situadas perpendicularmente al eje Z . En esta nueva posición, desde el punto de vista del eje de cotas, las curvas se presentan en 2 dimensiones con formas semejantes a elipses debido a la morfología somática próxima a la de una esfera (Figura 3.2). Para cada límite entre secciones se calculó la elipse que mejor aproximaba a la representación de la curva en el plano XY mediante mínimos cuadrados (Halir y Flusser, 1998) dando como resultado un ángulo o dirección del eje mayor respecto al eje de abscisas, el centroide de la elipse (dos dimensiones) y la longitud de sus dos ejes.

En el segundo paso se introdujo la tercera dimensión, la profundidad de la curva, que se aproximó mediante un polinomio de la forma

$$Z = \text{centroide} + \beta_1 X^2 + \beta_2 Y^2$$

donde Z es la tercera dimensión, X son las posiciones de los puntos de la elipse sobre el eje de abscisas, Y son las posiciones de los puntos de la elipse sobre el eje de ordenadas, centroide es el centroide de la elipse más el centroide de la curva sobre Z y β_1 y β_2 son los coeficientes del polinomio que mejor aproximaron la profundidad de la curva. Los valores de los coeficientes que mejor aproximaban la curva se calcularon mediante mínimos cuadrados.

3.2.2. Descriptores del soma

El proceso anterior basado en determinar el conjunto de parámetros de las elipses y coeficientes del polinomio que ajustaban las curvas de cada sección facilitó la

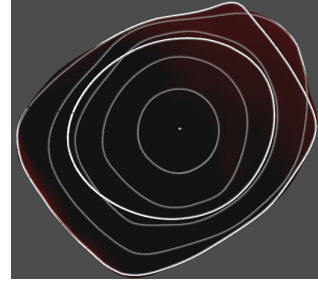


Figura 3.2: Curvas próximas a elipses: Se presenta un soma situado orientado sobre el eje de cotas y desde la vista perpendicular al plano XY . Desde este punto de vista las curvas presentan formas parecidas a elipses.

3.2 Caracterización basada en Reeb graph

definición de un conjunto de características locales y globales del soma. Los somas fueron aproximados por una secuencia de $K = 8$ secciones (2 puntos y 6 curvas) $S_i, i = 1, \dots, 8$. Para cada sección, las curvas definieron los límites superior T_i e inferior B_i sobre las que se asumió que T_i^R, B_i^R son los ejes mayores y T_i^r, B_i^r son los ejes menores de las elipses. De acuerdo al método utilizado para obtener las curvas, la superficie de la malla pudo ser interpretada como un continuo de puntos y en consecuencia se puede determinar que entre secciones adyacentes se cumple: $\forall, 1 < i < K + 1, B_i^R = T_{i-1}^R, B_i^r = T_{i-1}^r$. Se consideraron las siguientes medidas para caracterizar al soma:

- $|\vec{h}_i|, i = 1, \dots, K$: La altura relativa de cada sección. Es la longitud del vector \vec{h}_i definido por los centroides de dos elipses consecutivas.
- $r_i^h, i = 1, \dots, K - 1$: El cociente $\frac{B_{i+1}^R}{|h_i| + |h_{i+1}|}$ donde B_{i+1}^R es el eje mayor de la elipse que separa las secciones i -ésima y $(i + 1)$ -ésima y $|h_i|$ es la altura relativa de la sección i . Valores bajos para r_i^h denotan una sección elongada mientras que valores altos indican que la sección es achatada.
- $e_i, i = 1, \dots, K - 1$: Relación de aspecto de la i -ésima elipse o excentricidad ($B_{i+1}^r : B_{i+1}^R$) la cual se calcula a partir de $\frac{B_{i+1}^r}{B_{i+1}^R}$ donde B_{i+1}^r es el eje menor de la elipse que separa las secciones i -ésima e $i + 1$ -ésima y B_{i+1}^R es el eje mayor de la misma elipse. Cuando la relación de aspecto es igual a 1 la elipse es en realidad un círculo mientras que cuanto más pequeño se hace e más se aplana la elipse.
- $\alpha_i, i = 1, \dots, K - 1$: El ángulo entre los vectores \vec{h}_i y $(0, 0, 1)$. El primer vector se obtiene conectando los centroides i -ésimo e $i + 1$ -ésimo. El ángulo se obtiene a partir del arcoseno de $\cos \alpha_i = \frac{\vec{h}_i \cdot (0, 0, 1)}{|\vec{h}_i|}$.
- $\phi_i, i = 1, \dots, K - 1$: El ángulo azimuth o azimuthal se obtiene después de la transformación al sistema de coordenadas esféricas. Se toma el centroide de la elipse como el centro de la esfera. La punta del vector $\vec{h}_i \times (0, 0, 1)$ fue el punto proyectado a partir del cual el ángulo polar θ_i y el ángulo azimuthal ϕ_i fueron computados. El ángulo azimuth fue calculado de acuerdo a:

$$\phi = \begin{cases} \arctan\left(\frac{y}{x}\right) & x > 0, y > 0 \\ 2\pi + \arctan\left(\frac{y}{x}\right) & x > 0, y < 0 \\ \frac{\pi}{2} \operatorname{sgn}(y) & x = 0 \\ \pi + \arctan\left(\frac{y}{x}\right) & x < 0 \end{cases}$$

3. CARACTERIZACIÓN DEL SOMA

donde x, y y z son las coordenadas correspondientes a la punta del vector de $\vec{h}_i \times (0, 0, 1)$. Puesto que $(0, 0, 1)$ es el eje Z , θ_i es constante y no aporta información diferenciadora, no se incluyó dentro de la lista de características. El ángulo ϕ aporta información acerca de la dirección del vector \vec{h}_i sobre el plano XY .

- $\omega_i, i = 1, \dots, K-1$: Ángulo que forma el eje principal de la elipse obtenida tras orientar el soma sobre el eje Z y estudiar el soma desde el plano XY respecto al eje de abscisas. Esta variable aporta información acerca de la dirección de la curva sobre el plano XY .
- $\beta_{ij}, i = 1, \dots, K-1, j = 1, 2$: Coeficientes del polinomio $Z = \text{centroide} + \beta_{i1}X^2 + \beta_{i2}Y^2$ que representa la profundidad de la curva que separa cada una de las secciones.

Al tomar estas medidas asumiendo que $K = 8$ se obtuvieron un total de 50 descriptores para cada uno de los somas.

4

Modelización del soma

Mediante la neuromorfología teórica se pretende describir la forma y la conectividad del sistema nervioso. Uno de sus enfoques más extendidos consiste en analizar las características estructurales de componentes concretos, es decir, en realizar un estudio individualizado neurona por neurona de características morfológicas que permitan asociar las formas de los componentes a los roles que desempeñan. Este planteamiento se basa en el inductivismo, ya que trata de elaborar conclusiones generales a partir de observaciones individuales. El proceso inductivo es inherente a la observación y registro de los hechos, el análisis de los datos y la clasificación de los hechos. De acuerdo a este procedimiento, la primera fase fue abordada con el pre-procesamiento de la neurona y la transformación del soma a un conjunto de descriptores. En este capítulo se da continuidad al proceso con el análisis y clasificación de los datos mediante un modelo matemático basado en un nuevo subtipo de red bayesiana.

4.1. Definición del modelo

Para abordar el problema de la modelización de un conjunto de descriptores que representan una figura tridimensional se realizaron algunas asunciones con las que tratar el problema de manera más compacta.

- En primer lugar se aceptó la hipótesis de que el conjunto de datos se encuentra generado por una mixtura de redes bayesianas. Puesto que se persigue el análisis y clasificación de los datos, se motiva la aplicación de un algoritmo de clustering con el que explorar los datos y agrupar somas similares en el mismo grupo.

4. MODELIZACIÓN DEL SOMA

- En segundo lugar se hipotetizó que las variables de la red Bayesiana siguen una distribución Gaussiana dado que se trata de una distribución de probabilidad que permite obtener buenas aproximaciones.
- En tercer lugar se asumió un modelo Markoviano espacial de orden 1 entre las secciones de la figura, esto es, dada la sección previa o inferior, la sección actual es independiente del resto de las anteriores. De esta forma el número de dependencias entre las variables del modelo se redujo, condición restrictiva necesaria dado el reducido número de observaciones en contraste con los 50 descriptores. Además, esta estructura concuerda con el patrón seguido en la simulación para la representación tridimensional de un soma a partir de los descriptores en el que cada sección se construye a partir de la anterior.
- Una última asunción que se definió a partir del modelo markoviano consistió en mantener una estructura fija para la red bayesiana de manera que los arcos entre nodos dentro de una sección y los arcos de una sección a la siguiente (arcos de transición) se mantuvieran en todos los niveles de la red bayesiana (Figura 4.1B). La red bayesiana resultante comparte características con la red bayesiana dinámica a menudo aplicada para modelar datos secuenciales o series temporales; de ahí que el modelo propuesto sea considerado una adaptación de los modelos temporales a los modelos espaciales. En la Figura 4.1 se muestra un ejemplo.

4.1.1. Red bayesiana gaussiana lineal

Cuando se trata de modelar variables continuas la forma paramétrica más utilizada para funciones de densidad continua es la distribución gaussiana. Para explicar el problema derivado de la representación de la dependencia entre dos variables continuas gaussianas se puede considerar una situación entre dos variables aleatorias tal que X es el padre de Y . Una solución simple al problema consiste en modelar la distribución de Y como una gaussiana, cuyos parámetros dependen del valor de X que puede asumir infinitos valores. En estas circunstancias la media de Y puede convertirse en una función lineal dependiente de X , mientras que la varianza de Y no depende de X . De esta forma se obtendría, por ejemplo

$$p(Y|x) = \mathcal{N}(-2x + 0,9; 1).$$

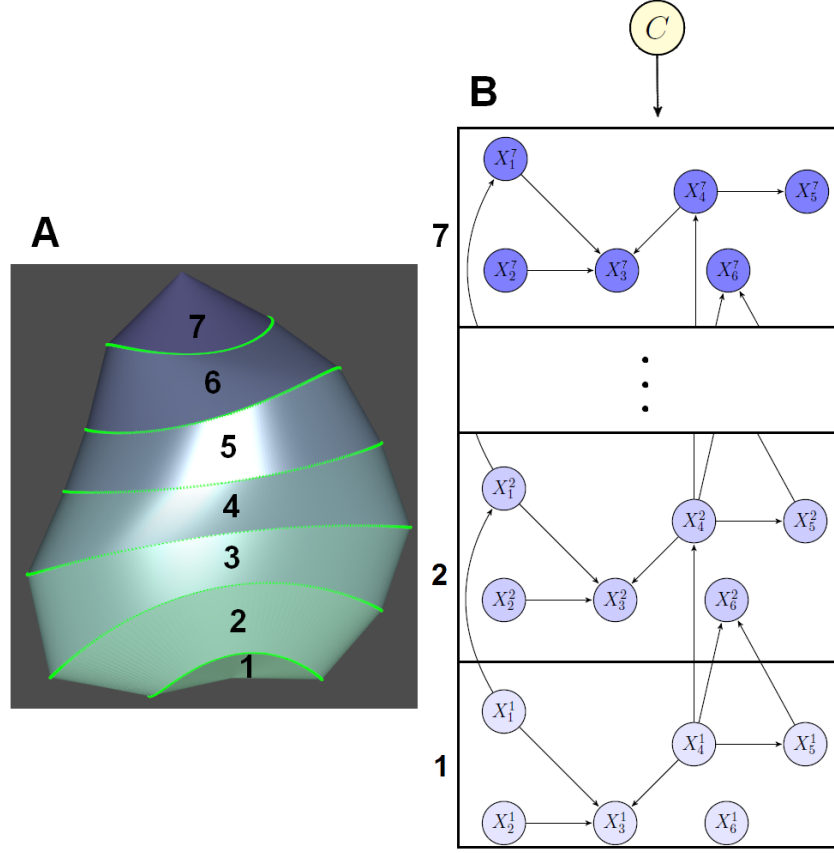


Figura 4.1: Red bayesiana dinámica espacial: (A) Imagen de un soma con cada una de sus secciones numeradas en orden ascendente. (B) Ejemplo de red bayesiana dinámica espacial para el soma de la imagen A. Se puede observar cómo las relaciones de dependencia inter-sección se repiten para cada una de las secciones al igual que las relaciones entre distintas secciones, siguiendo un modelo markoviano. En la parte superior se sitúa la variable latente C que categoriza al soma. Para simplificar la representación no se incluyeron los arcos de dependencia desde C hacia todas las variables.

Este tipo de dependencia se denomina modelo lineal gaussiano y puede ser extendido para más variables padres de forma directa mediante

$$p(Y|x_1, x_2, \dots, x_k) \sim \mathcal{N}(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k; \sigma^2).$$

De acuerdo a las definiciones anteriores, una red bayesiana gaussiana es una red bayesiana en la que todas sus variables son continuas y donde todas sus distribuciones de probabilidad condicionada son gaussianas lineales. Una red gaussiana lineal siempre define una distribución conjunta multivariante gaussiana de manera que si

4. MODELIZACIÓN DEL SOMA

Y es una gaussiana lineal cuyos padres son X_1, \dots, X_k :

$$p(Y|\mathbf{x}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}; \sigma^2)$$

Si se asume que X_1, \dots, X_k son gaussianas conjuntas con distribución $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces, la distribución de Y es una distribución normal $p(Y) = \mathcal{N}(\mu_Y; \sigma_Y^2)$ donde:

$$\mu_Y = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{\mu}$$

$$\sigma_Y^2 = \sigma^2 + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$$

4.2. Aprendizaje de la estructura de la red

La estructura de una red bayesiana puede ser utilizada para descubrir conocimiento mediante el estudio de sus dependencias, con las que se pueden revelar las relaciones de dependencia entre las variables del dominio. Si bien es cierto que mediante tests de independencia condicional también se puede obtener las dependencias entre variables, la estructura de la red bayesiana distingue entre dependencias directas e indirectas.

Los algoritmos para el aprendizaje de estructuras de redes bayesianas se dividen en tres grandes grupos:

- Un enfoque utiliza el aprendizaje de estructuras basado en restricciones que considera la red bayesiana como una representación de independencias. Este grupo de algoritmos comprueba la dependencia condicional y la independencia en los datos con el objetivo de encontrar una red (o un conjunto de clases de redes equivalentes) que explique lo mejor posible las dependencias e independencias obtenidas en la base de datos.
- El segundo enfoque es denominado aprendizaje de estructura basado en puntuación. Este conjunto de métodos consideran a la red bayesiana como una especificación de un modelo estadístico para posteriormente abordar el aprendizaje como un problema de selección de modelos. Todos estos algoritmos operan bajo el mismo principio: Se define un espacio de hipótesis o modelos potenciales (modelos que se van a considerar como estructuras de la red) y una función de puntuación que mide cómo de bien se ajusta el modelo a los datos. La cardinalidad de este espacio es superexponencial por lo que se trata de un problema \mathcal{NP} -duro lo que motiva el uso de heurísticas. Se escogió un

modelo de este conjunto debido a que es menos sensible a fallos individuales y suele adquirir un compromiso entre la dependencia de las variables y el coste de añadir un nuevo arco.

- El tercer enfoque no intenta aprender una única estructura sino que en su lugar genera un conjunto de estructuras posibles. Con este método denominado promedio de modelos bayesianos se pretende hacer una media con las predicciones de todas las posibles estructuras. Dado que el número de estructuras es inmensa realizar esta labor puede resultar imposible aunque para ciertos problemas se puede lograr de manera eficiente.

El método escogido para el aprendizaje de la estructura fue *Hill-Climbing* (HC). La elección de este algoritmo sobre el resto se debió principalmente al número limitado de observaciones de las que se dispuso (40 somas) respecto al número de descriptores (más de 50) ya que HC presentó menor tendencia a incluir arcos en la estructura. De esta forma se trató de evitar problemas derivados del sobreajuste.

HC comienza la búsqueda a partir de una red vacía, completa o iniciada aleatoriamente. A continuación estima los parámetros de las distribuciones de probabilidad locales dada la estructura de la red bayesiana. Típicamente se utiliza un estimador de máxima verosimilitud. El bucle principal del algoritmo consiste en intentar introducir, eliminar o revertir cada uno de los arcos posibles o existentes de manera que la red aumente la puntuación lo máximo posible en cada iteración. El proceso se detiene cuando no existen cambios que incrementen el valor de la puntuación. El algoritmo no garantiza encontrar el máximo global por lo que a menudo se realizan cambios o perturbaciones para aumentar la probabilidad de alcanzarlo. Otra posibilidad es reiniciar múltiples veces desde puntos aleatorios (redes iniciales) en el espacio. Se aprovechó la implementación provista en el paquete de R *bnlearn* (Nagarajan et al., 2013; Scutari, 2010).

4.2.1. Red bayesiana dinámica espacial

Con el objetivo de lograr una estructura similar a la de una red bayesiana dinámica (Figura 4.1) se siguió un procedimiento similar al aplicado en el aprendizaje de sus homólogas temporales. Puesto que los arcos internos, es decir, dentro de una sección y los arcos de transición se repiten para todas las secciones, la estructura debe de ser lo más general posible. Para ello se generó un nuevo conjunto de datos como se muestra en la Tabla 4.1

4. MODELIZACIÓN DEL SOMA

	$ \vec{h}_i $	r_i^h	...	β_{i2}	$ \vec{h}_{i+1} $	r_{i+1}^h	...	$\beta_{(i+1)2}$
Observación 1	-	-	-	-	-	-	-	-
Observación 2	-	-	-	-	-	-	-	-
Observación 3	-	-	-	-	-	-	-	-
...	-	-	-	-	-	-	-	-
Observación $(K - 1) * N$	-	-	-	-	-	-	-	-

Tabla 4.1: Conjunto de datos para el aprendizaje de la estructura de la red:

Por un lado, las variables de las secciones $i = 1, \dots, K - 1$ se agruparon para aprender la estructura interna general de las secciones. Por otro lado se agruparon las variables de las secciones $i + 1 = 2, \dots, K$ que, junto con las observaciones de las secciones i , se utilizaron para aprender los arcos de transición entre las secciones i e $i + 1$. Al agrupar las variables se obtienen un total de $(K - 1) * N$ donde N es el número de somas. Se inhabilitó la inclusión de arcos de transición de secciones superiores a inferiores.

4.2.2. Procedimiento bootstrap para aprendizaje robusto

En problemas complejos de análisis de datos es necesario no solo que se induzca la red bayesiana a partir de los datos sino que también se provean medidas de la confianza acerca de las características de la red. Por ejemplo, se podría discutir si está garantizada la existencia de un arco entre dos nodos, si el manto de Markov de un nodo es robusto o si se puede conocer cierta información acerca del orden de las variables. Se desearía poder responder a estas cuestiones aún cuando la cantidad de datos no es suficiente (Friedman et al., 1999). Un método para resolver este problema es utilizar *bootstrap* (Efron, 1979; Efron y Tibshirani, 1994).

Sea G la estructura de una red. Una propiedad de interés en esta estructura puede ser la existencia de $X \rightarrow Y$ en G o que X preceda a Y . En general, estas propiedades puede ser tratadas como funciones f a partir de la estructura de la red con el conjunto $\{0, 1\}$, donde 1 es que se cumple la propiedad y 0 que no se cumple.

Se asume que se dispone de un conjunto de datos con N observaciones $D = x^{(1)}, \dots, x^{(N)}$ que asigna valores a X_1, \dots, X_n y, además, que estas asignaciones fueron muestreadas independientemente a partir de una red probabilística B con estructura G . Siendo $\hat{G}(D)$ la estructura de la red devuelta por el algoritmo de inducción al que se le introdujo como entrada D , para cualquier propiedad f se considera la siguiente cantidad

$$p_N(f) = Pr\{f(\hat{G}(D)) = 1 \mid |D| = N\},$$

que es la probabilidad (p_N) de inducir una red con la característica f entre todos los posibles conjuntos de datos ($|D|$) de tamaño N que pueden ser muestreados a partir de B . Si el proceso de inducción es consistente, se puede esperar que según N se haga más grande, $p_N(f)$ converja a $f(G)$. Esto es, la confianza en f será cercana a 1 si se mantiene en G , y cercana a 0 si no lo es.

Un método que permite estimar $p_N(f)$ dado un único conjunto de observaciones D de tamaño N es el método no paramétrico bootstrap. La idea subyacente tras el método es que se puede confiar más en las propiedades que se siguen induciendo cuando los datos son “distorsionados”. La cuestión ahora es cómo distorsionar los datos de manera que aun así se mantengan las propiedades estadísticas generales del conjunto de datos. En la versión no paramétrica de bootstrap tales distorsiones se generan re-muestreando a partir de un conjunto de datos dado. A continuación se estima la confianza en una propiedad mediante el examen de cuántas veces aparece la propiedad en los conjuntos de datos distorsionados. El método no paramétrico de bootstrap funciona del siguiente modo:

- Para $i = 1, 2, \dots, m$
 - Remuestrear, con reemplazamiento, N instancias a partir de D . Denotar con D_i cada uno de los conjuntos de datos resultantes.
 - Aplicar el procedimiento de aprendizaje en D_i (en este proyecto HC) para inducir una estructura de red $\hat{G}_i = \hat{G}(D_i)$.
- Para cada propiedad de interés definir

$$p_N(f) = \frac{1}{m} \sum_{i=1}^m f(\hat{G}_i).$$

Como resultado del proceso anterior se obtiene un valor entre 0 y 1 que determina la confianza de esa propiedad en la estructura de la red. A partir de un umbral de confianza se determinan qué propiedades prevalecen en la estructura y cuáles no son incluidas.

En este proyecto se aplicó *bootstrap* para determinar el nivel de confianza en los arcos de \hat{G} . Para cada uno de los 200 conjuntos de datos generados se obtuvo una estructura mediante el método HC y se contabilizó la existencia de arcos entre descriptores teniendo en cuenta su dirección. Finalmente, se impuso un umbral de confianza subjetivo del 80 % sobre el resultado de bootstrap que determinó qué arcos debían delimitar la estructura final de la red (Figura 4.2). Una vez se obtuvo la

4. MODELIZACIÓN DEL SOMA

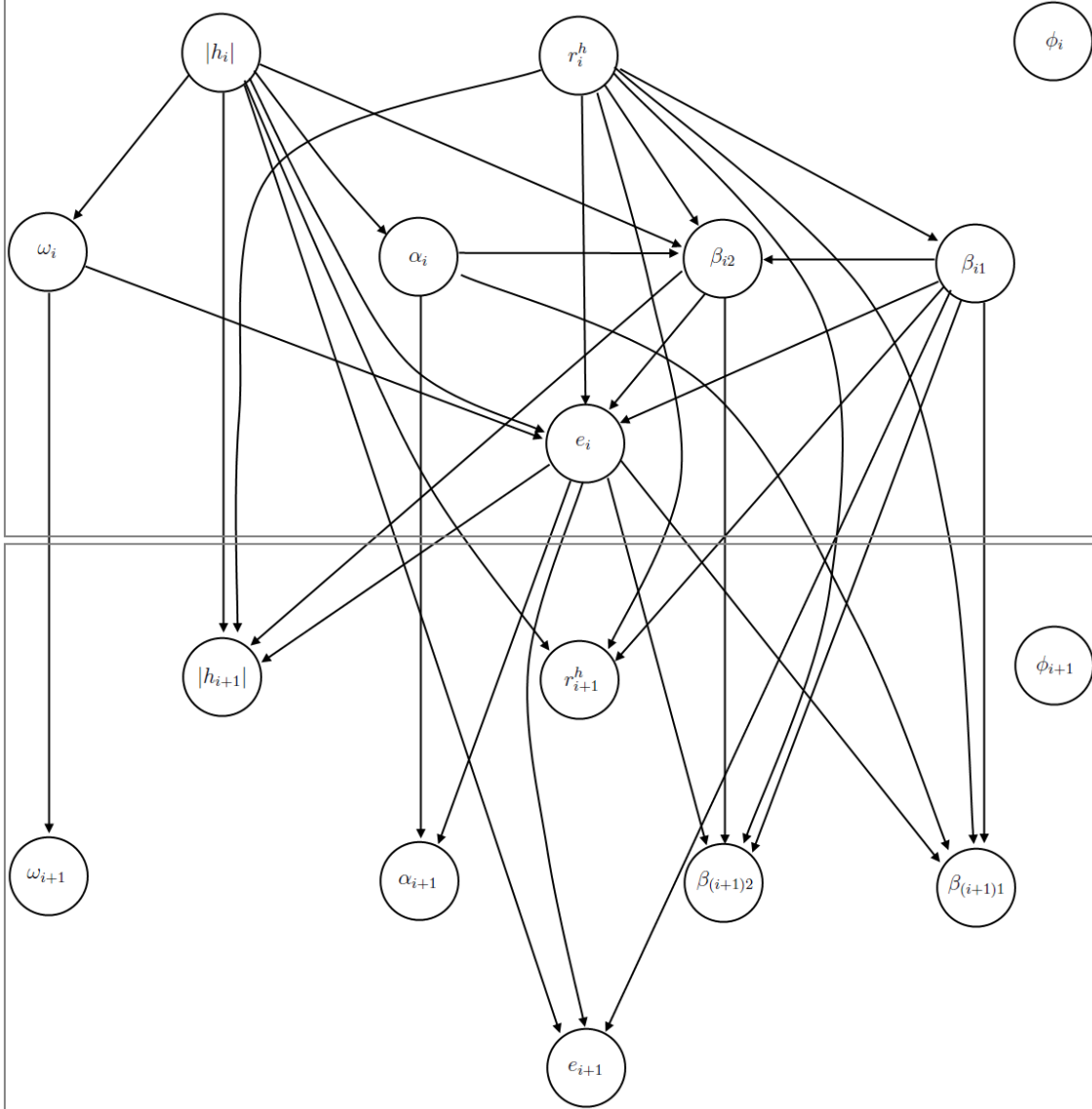


Figura 4.2: Estructura de la red bayesiana en las secciones i e $i+1$. Se muestra la estructura de la red bayesiana con los arcos internos de una sección, denotados por los arcos que unen variables con el subíndice i , y los arcos de transición que unen los arcos con subíndice i con los de subíndice $i+1$.

estructura de red bayesiana para las secciones i y $i+1$, la estructura se extendió para todas las secciones.

4.3. Clustering probabilístico de somas basado en una red bayesiana dinámica espacial

En un modelo basado en clustering mediante redes bayesianas cada variable o nodo de la red modela una distribución de probabilidad y el conjunto de todas las distribuciones se combina para dar lugar a un modelo finito de mixturas M . Para obtener el modelo de mixtura de redes bayesianas se estima el conjunto completo de parámetros mediante la estimación de máxima verosimilitud (EMV), esto es, se trata de encontrar el conjunto de parámetros θ que maximiza la verosimilitud de los datos observados $\arg \max_{\theta} P_{\theta}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$ donde $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$ son los datos. Habitualmente se utiliza la log-verosimilitud en lugar de la verosimilitud porque facilita los cálculos. El problema que surge cuando se intenta maximizar la log-verosimilitud para una mixtura es que la fórmula matemática no es cerrada:

$$\ln P_{\theta}(\mathbf{D}|\mathbf{M}) = \sum_{n=1}^N \ln \left[\sum_{k=1}^K \pi_k p(\mathbf{x}^{(n)}|M_k) \right],$$

donde \mathbf{D} son los datos, $\mathbf{M} = (G, \theta)$ representa el modelo finito de mixturas, M_k denota la componente de la mixtura, la estructura y parámetros de la clase k tal que $\sum_i^K = 1$, $\pi_k \in (0, 1]$, N el número de observaciones, K el número de clases, π_k la probabilidad a priori de la clase k . Calcular un máximo puede resultar muy complejo debido al logaritmo de la suma. Una solución para abordar este problema fue propuesta en Dempster et al. (1977) que permite lograr una aproximación utilizando un procedimiento iterativo de dos pasos denominado algoritmo esperanza-maximización (EM).

4.3.1. Algoritmo Esperanza-Maximización

4.3.1.1. Paso de esperanza

EM comienza inicializando los parámetros θ , bien condicionándolos al conocimiento que se tenga acerca del problema, bien utilizando algún criterio o algoritmo de inicialización o bien aleatoriamente. A continuación se evalúa la log-verosimilitud del modelo \mathbf{M} con los parámetros generados. El siguiente paso es el de esperanza en

4. MODELIZACIÓN DEL SOMA

el que se calcula la responsabilidad del modelo k para la n -ésima observación, esto es, se evalúa la probabilidad a posteriori de que la n -ésima observación pertenezca al cluster k :

$$\gamma(k_n) = \frac{\pi_k p(\mathbf{x}_n | M_k)}{\sum_{j=1}^K \pi_j p(\mathbf{x}_n | M_j)}.$$

Subdesbordamiento en el cálculo de responsabilidades A nivel computacional, la multiplicación de probabilidades puede dar lugar a números muy pequeños que el ordenador redondee a 0, es decir, se puede producir subdesbordamiento y en consecuencia divisiones entre 0. A medida que el número de variables aumenta es más fácil que acabe ocurriendo. Esta situación se dió en el cálculo de $\gamma(k_n)$ y, para solucionar el problema, se hizo uso del método *log-sum-exp*. Esta técnica realiza una operación de desplazamiento para poder llevar a cabo el cálculo que produce el subdesbordamiento utilizando propiedades del logaritmo. Posteriormente se elimina el desplazamiento y se calcula de nuevo el exponente. El método quedó definido del siguiente modo:

$$\gamma(k_n) = \exp(\log(\pi_k) + \log(p(\mathbf{x}_n | M_k)) - b - \log(\sum_{j=1}^K \exp(\log(\pi_j) + \log(p(\mathbf{x}_n | M_j)) - b))),$$

donde $b = \max_k p(\mathbf{x}_n | M_k)$.

4.3.1.2. Paso de maximización

El paso de maximización re-estima los parámetros utilizando las responsabilidades calculadas en el paso de esperanza. Así, los parámetros θ son optimizados en sus nuevas versiones. Dado que el proceso de clustering se desarrolló sobre una red bayesiana los parámetros se calcularon localmente para cada nodo. Al tratarse de una red bayesiana gaussiana lineal, para calcular los parámetros de un nodo descendiente de otro primero se necesitan los parámetros de los ancestros con lo que para la estimación de parámetros se sigue una búsqueda en anchura sobre el grafo de la red. La estimación de parámetros es como sigue:

- Para un nodo sin padres la función de densidad es $\mathcal{N}(X|\mu, \sigma^2)$ por lo que sus

4.3 Clustering probabilístico de somas basado en una red bayesiana dinámica espacial

parámetros de máxima verosimilitud se calculan a partir de:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(k_n) x_n$$

$$\sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(k_n) (x_n - \mu_k^{new})^2$$

- Para un nodo con l padres la función de densidad es $\mathcal{N}(X|\beta_{0k} + \beta_{1k}u_1 + \beta_{2k}u_2 + \dots + \beta_{lk}u_l, \sigma^2)$, donde u_t son los datos correspondientes al padre t y β_{tk} es el coeficiente para el padre t y la clase k . Para estas variables en lugar de estimar un único valor μ , se deben de hallar los estimadores de máxima verosimilitud para los l coeficientes. Como resultado se obtiene un sistema de ecuaciones de la siguiente forma:

$$\begin{array}{ccccccc} \beta_{0k} \sum_{n=1}^N \gamma(k_n) + \beta_{1k} \sum_{n=1}^N \gamma(k_n) u_{1n} & + \dots & + \beta_{lk} \sum_{n=1}^N \gamma(k_n) u_{ln} & = & \sum_{n=1}^N \gamma(k_n) x_n \\ \beta_{0k} \sum_{n=1}^N \gamma(k_n) u_{1n} + \beta_{1k} \sum_{n=1}^N \gamma(k_n) (u_{1n})^2 & + \dots & + \beta_{lk} \sum_{n=1}^N \gamma(k_n) u_{ln} u_{1n} & = & \sum_{n=1}^N \gamma(k_n) x_n u_{1n} \\ \vdots & & \vdots & & \\ \beta_{0k} \sum_{n=1}^N \gamma(k_n) u_{ln} + \beta_{1k} \sum_{n=1}^N \gamma(k_n) u_{1n} u_{ln} & + \dots & + \beta_{lk} \sum_{n=1}^N \gamma(k_n) (u_{ln})^2 & = & \sum_{n=1}^N \gamma(k_n) x_n u_{ln} \end{array}$$

Con los coeficientes se calcula el parámetro σ^2 del siguiente modo:

$$\sigma_k^{new} = \frac{\sum_{n=1}^N \gamma(k_n) (x_n - \beta_{0k} - \beta_{1k}u_{1n} - \dots - \beta_{lk}u_{ln})^2}{\sum_{n=1}^N \gamma(k_n)}.$$

Finalmente se calcula $\pi_k^{new} = \frac{N_k}{N}$ donde $N_k = \sum_{n=1}^N \gamma(k_n)$.

A continuación, se evalúa la log-verosimilitud

$$\ln P(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\theta}^{new}) = \sum_{n=1}^N \ln \left[\sum_{k=1}^K \pi_k^{new} p(\mathbf{x}_n | M_k^{new}) \right].$$

El algoritmo EM itera hasta que o bien los parámetros o la log-verosimilitud convergen. Una vez finalizó EM y se obtuvieron los parámetros para el modelo de mixturas, se devolvió la distribución de probabilidad de pertenencia a cada clase para cada soma. Con ello se alcanzó un “soft” clustering, es decir, los somas no pertenecían a una única clase sino que pertenecían a distintas clases con una determinada probabilidad.

4. MODELIZACIÓN DEL SOMA

4.3.2. Selección del número de clases

Puesto que se desconoce el número de clases K que conforman la mixtura de redes bayesianas y, de hecho, es parte de la información que se pretendió revelar mediante la aplicación del algoritmo, K se escogió de acuerdo a una puntuación. La selección de K se basó en un proceso iterativo que consistió en ejecutar el algoritmo de clustering variando el número de clases y calculando para cada iteración la puntuación. Al finalizar se escogió el número de clases que obtuvo la mejor puntuación. Una selección natural como mecanismo de puntuación podría ser la log-verosimilitud ya que este valor evalúa lo bien que se ajusta el modelo a los datos. Sin embargo, este valor provoca que el modelo se pueda volver excesivamente complejo debido a que un mayor número de parámetros permite ajustar con más precisión un conjunto de datos de entrenamiento lo que a la larga produce sobreajuste u “overfitting”. Desarrollar un modelo que sufra sobreajuste implica que el modelo pierda poder predictivo debido a que ajusta en exceso pequeñas fluctuaciones en los datos. A partir de este problema se establece que es necesario un equilibrio entre el número de parámetros y el ajuste de los datos al modelo.

Con el fin de definir un criterio para la selección de modelos entre un conjunto finito de modelos se utilizó el denominado *Bayesian Information Criterion* (BIC) también conocido como criterio de Schwarz (Schwarz, 1978). Se trata de una puntuación que añade una penalización a la log-verosimilitud basada en el número de parámetros v_M del modelo. Esta forma, puede ser utilizada para elegir el modelo de clustering (parametrización y número de clases) intentando evitar la selección de modelos complejos. El cálculo de BIC se realizó mediante:

$$BIC_M = 2 \ln P(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\theta}, \mathbf{M}) - v_M \ln(N)$$

donde $\boldsymbol{\theta}$ son los parámetros estimados mediante EM, \mathbf{M} es el modelo de la red bayesiana, $\mathbf{x}_1, \dots, \mathbf{x}_N$ son los datos observados y N el tamaño de la muestra.

Como paso adicional también se calculó *Akaike Information Criterion* (AIC), otra puntuación para modelos pero que añade una penalización menor que BIC a la log-verosimilitud. AIC fue calculado mediante:

$$AIC_M = 2 \ln P(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\theta}, \mathbf{M}) - 2v_M.$$

El resultado de ambos criterios de selección de modelo se muestran en la Figura 4.3:

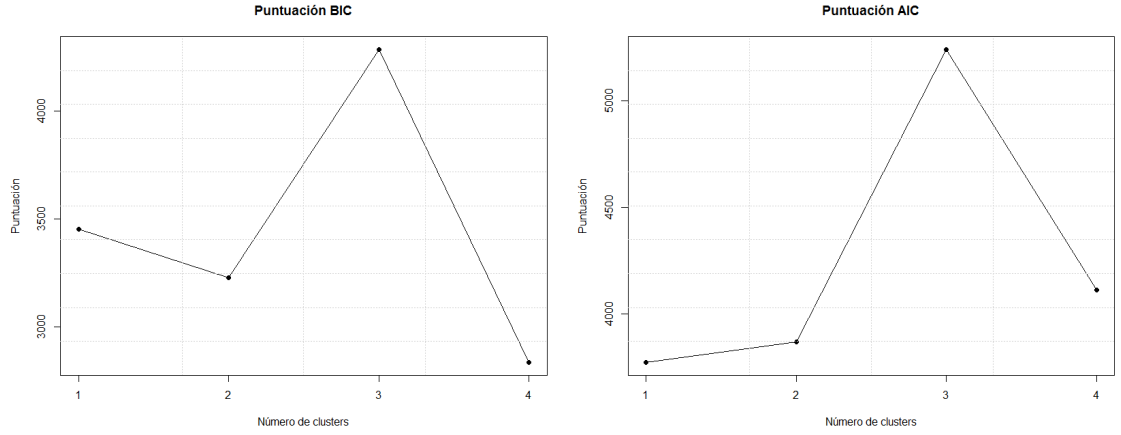


Figura 4.3: Selección del número de clusters. Se observa que para ambos criterios de selección la mejor puntuación se obtuvo claramente para 3 clusters. Además, se demuestra claramente la diferencia en la penalización de ambos criterios, mientras que en BIC el segundo mejor modelo se representa con 1 cluster (modelo más simple) y el peor con 4 (modelo más complejo), con AIC ocurre justamente lo contrario.

4.4. Resultados

4.4.1. Relaciones de dependencia entre los descriptores

Con la especificación de una red bayesiana para los descriptores del soma se obtuvo una estructura que representó las dependencias entre variables. Esta información se muestra en forma de grafo acíclico dirigido en la Figura 4.2. La estructura se extiende desde la sección $i = 1$ hasta $i + 1 = K$. Con el análisis de los arcos entre descriptores de una misma sección, los arcos de transición y los coeficientes de los padres de cada nodo se adquirió una comprensión más profunda del dominio.

4.4.1.1. Dependencia inter-sección

Con el estudio de los arcos entre los nodos de una misma sección se observaron varios resultados de interés.

En la Figura 4.2 se muestra que los ángulos entre secciones α_i y los ángulos de las elipses ω_i dependen directamente de la altura $|h_i|$. Este patrón se presentó debido al método basado en rayos para definir las curvas que diferenciaron las secciones que provocó que en los extremos del soma las curvas se encontrasen más próximas entre ellas y en su mitad más alejadas, es decir, que la altura entre secciones es menor en los extremos y mayor en el centro. Por lo tanto se originaron dependencias en la

4. MODELIZACIÓN DEL SOMA

estructura de la red que indicaron que la curvatura del soma es distinta entre el centro y la parte media del soma y que la máxima anchura en las secciones intermedias difiere de las de los extremos, es decir, que la morfología no es homogénea entre secciones.

También se observó dependencia entre el achatamiento de la sección r_i^h y los coeficientes de la curva entre secciones β_{i1} y β_{i2} . Esta relación cobra sentido si se considera que r_i^h relaciona el radio mayor de elipse que aproxima la curva entre secciones con la altura de la sección superior e inferior. Por lo tanto aporta información de la curva del plano 2-dimensional XY sobre el que interactúan los coeficientes β_{ij} . r_i^h aporta información acerca del rango de valores que asumen las incógnitas del polinomio que dictamina la curvatura de la curva entre secciones y, dado que el cálculo de los coeficientes depende directamente de esos datos, se identifica una relación. De nuevo, debido al método utilizado para definir las curvas entre secciones basado en el trazado de rayos, las secciones situadas en los extremos presentan curvas cuyo radio mayor es más pequeño que para las curvas intermedias. Las curvas intermedias suelen presentar menor curvatura que las de los extremos como se puede observar en la Figura 3.1. Por lo tanto, a partir del achatamiento de la sección se infiere el grado de curvatura entre secciones.

Otro patrón que se muestra en la red bayesiana es la dependencia de los coeficientes de la curva entre secciones y la excentricidad, es decir, el arco $\beta_{ij} \rightarrow e_i$. En esta ocasión la variable dependiente relaciona los ejes de la elipse que aproxima la curva sobre el plano XY . En los extremos del soma es en la posición en la que los ejes menor y mayor toman los valores más pequeños y donde se situaron las mayores pendientes, por lo tanto los coeficientes β_{ij} tomaron valores más extremos. En las elipses centrales del soma apenas se encuentra pendiente y es la posición en la que el soma es más ancho y más grueso lo que implicó ejes más grandes y coeficientes más pequeños.

4.4.1.2. Dependencia entre secciones

Al analizar los arcos de transición entre secciones se observó que todos los descriptores $i+1$ dependieron de su descriptor i menos ϕ_i y ϕ_{i+1} que son independientes. Este hecho pudo deberse a que al tratar de modelar ambas variables mediante una distribución gaussiana no se aproxime adecuadamente la distribución subyacente angular. Es por ello que una futura tarea a realizar es tratar de modelarlas mediante una distribución von Mises o una distribución Wrapper Normal con el fin de obtener

un modelo que se ajuste mejor a ambas variables y que testee si ambas variables son dependientes o no lo son en un espacio angular.

Otra característica destacable que se encuentra representada en el grafo es que las variables que identifican a las características angulares únicamente dependen de si mismas y la altura. Dado este patrón de cuasi-independencia entre variables angulares y lineales que presentó la estructura de la red bayesiana se motiva la aplicación de un método de modelado basado en una distribución von Mises o bien en una distribución Wrapper Normal que permita ajustar mejor las relaciones entre variables angulares.

4.4.2. Distribución de las clases

4.4.2.1. Respecto a la localización cerebral

Con el objetivo de obtener una comprensión más profunda acerca de las clases de somas definidas por el clustering se estudió cómo varía su distribución en función de la capa de la cual fueron extraídos. Para ello se representó la proporción de clases para cada una de las localizaciones cerebrales como se muestra en la Figura 4.4.

En la gráfica de barras se puede observar que la clase 1 se encuentra especialmente localizada en la capa frontal mientras que la clase 2 se presenta en la capa cingular. La clase 3, que es la que cuenta con una menor representación, destaca en la capa temporal. A partir de la gráfica se puede dilucidar una relación entre las clases encontradas por el algoritmo de clustering y la capa cerebral que ocupaban. Debido al reducido número de observaciones no se pudo realizar un test de hipótesis que permitiese profundizar en el análisis de este patrón. Por lo tanto queda pendiente como parte del trabajo futuro demostrar o rechazar esta relación a la espera de más datos.

4.4.2.2. Respecto a la similitud morfológica

En el capítulo anterior se extendió la definición de esfericidad para cuerpos platónicos. La esfericidad, como se indicó, consiste en comparar el área de la superficie de un objeto tridimensional con el área de la superficie de una esfera del mismo volumen. El mismo cálculo se realizó para la pirámide, el cubo, el octaedro, el dodecaedro y el icosaedro.

Una relación exacta entre el objeto estudiado y el cuerpo platónico se obtiene cuando $\frac{A_o}{A_c} = 1$ donde A_o es el área del objeto y A_c es el área del cuerpo platónico, es

4. MODELIZACIÓN DEL SOMA

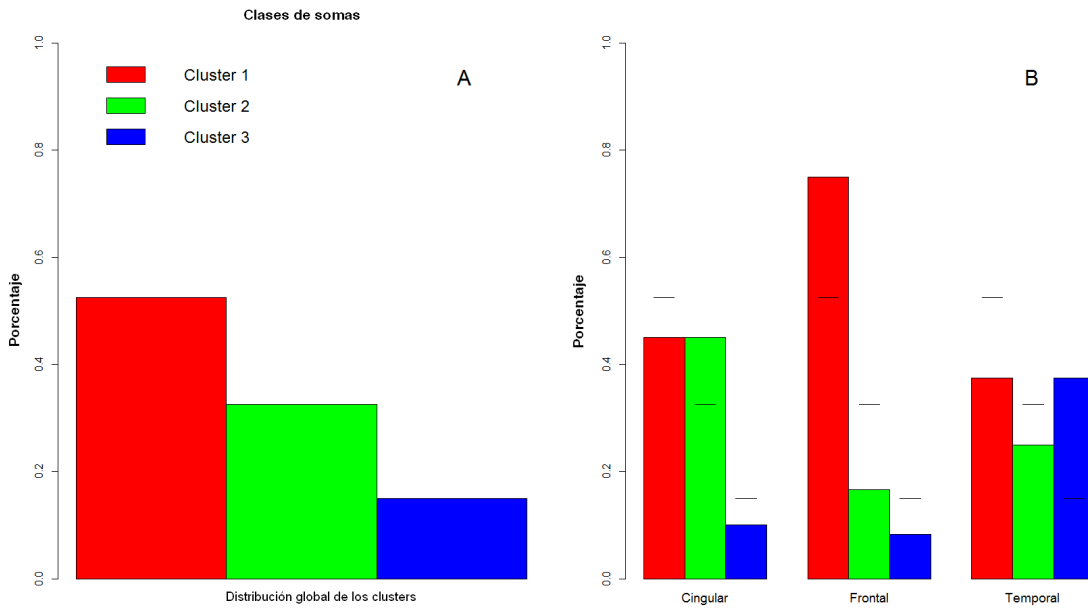


Figura 4.4: Distribución de los somas de acuerdo a la localización cerebral de la que se extrajeron. Diagrama de barras mostrando la distribución de los somas. (A) Distribución de los somas en las tres clases. (B) Distribución de probabilidad de las clases para los somas cingulares (izquierda), frontales (centro) y temporales (derecha). Las líneas horizontales en (B) denotan la proporción de las clases en (A).

decir, en caso de cumplirse la igualdad el objeto y el poliedro con el que se compara comparten el mismo volumen y el mismo área de la superficie. A partir de esta relación es posible definir un método de clasificación de manera que la morfología del soma se identifique con el cuerpo platónico con el que mantenga la relación más próxima a 1. A partir de esta nueva clasificación y de la ofrecida por el clustering en la red bayesiana dinámica espacial se analiza la distribución de los somas. El resultado se muestra en la Figura 4.5.

En la representación gráfica se aprecia que ninguna morfología se asemeja a una pirámide. Este dato contrasta con el nombre de este tipo de neuronas. Sin embargo, si se considera que la denominación proviene de observaciones realizadas mediante un microscopio de menor resolución que los actuales y únicamente en 2 dimensiones se puede justificar esta discrepancia. La pirámide es un cuerpo mucho más simple que un dodecaedro o un icosaedro por lo que estos dos se ajustan mejor a figuras tridimensionales complejas. De acuerdo a esta hipótesis, es razonable que tanto para la esfericidad como para la cubicidad se presente un único soma y que se encuentre una mayor representación en morfologías más complejas. Si se acepta esta hipótesis

de complejidad y de acuerdo a la gráfica presentada, la clase 2 presentaría una morfología más sencilla por encontrarse especialmente representada en la octaedricidad considerando su incremento respecto a la distribución global. La clase 1 destaca en la dodecaedricidad por lo tanto fue algo más compleja que la clase 2. Finalmente, de acuerdo a este criterio, la clase 3 fue la más compleja de todas.

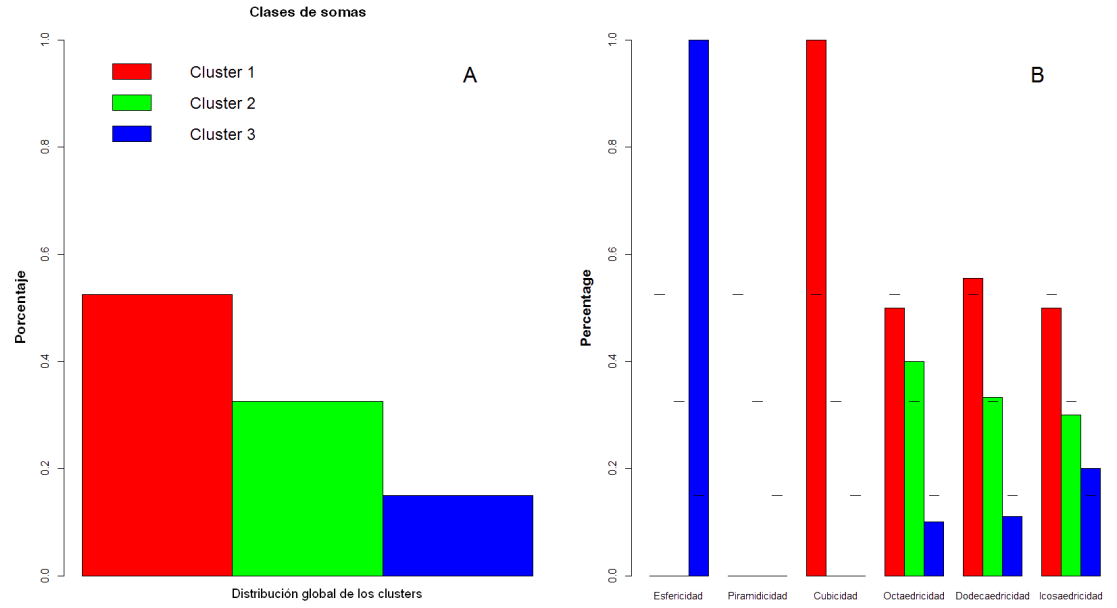


Figura 4.5: Distribución de los somas de acuerdo a la similitud morfológica. Diagrama de barras que representa la distribución de los somas con respecto al poliedro platónico al que más se asemejan, es decir, aquel cuyo $\frac{A_o}{A_c}$ más se aproxima a 1. (A) Distribución de los somas en las tres clases. (B) Distribución de probabilidad de las clases para los somas, de izquierda a derecha: esféricos, piramidales, cúbicos, octaédricos, dodecaédricos e icosaédricos. Las líneas horizontales en (B) denotan la proporción de las clases en (A).

4. MODELIZACIÓN DEL SOMA

5

Simulación de somas e interpretación del modelo

Por simulación se entiende la reproducción de la morfología cerebral así como la dotación de funcionalidad a cada uno de los componentes neuronales. A partir de la caracterización del soma y el desarrollo de la red bayesiana dinámica espacial fue posible simular la morfología de nuevas somas. Al comienzo de este capítulo se presenta el algoritmo de simulación y posteriormente se identifican las características discriminatorias de cada una de las clases de somas encontradas por la red bayesiana dinámica espacial.

5.1. Simulación

El proceso de simulación consistió en el muestreo de las características somáticas definidas en la red bayesiana dinámica espacial con el fin de obtener nuevas somas virtualizados. Se distinguen dos fases fundamentales. En el primer paso se muestrean nuevas observaciones de acuerdo a la distribución de probabilidad representada por la red bayesiana. En segundo lugar para cada una de las nuevas somas simuladas se genera una representación 3D a partir de las propiedades del soma muestreado.

5.1.1. Muestro de somas de cada clase

El método de clustering propuesto en el capítulo anterior dio lugar a una distribución de probabilidad multivariante representada por una red bayesiana dinámica espacial. La estructura de la red reproducida mediante un grafo acíclico dirigido identificó las relaciones de dependencia entre las propiedades somáticas mientras que el

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

algoritmo EM optimizó los parámetros para cada una de las clases. En disposición de la información estructural y de los parámetros de cada clase se muestrearon nuevos cuerpos somáticos virtualizados. Esta labor consistió en muestrear valores para las características de los nuevos somas de acuerdo a la distribución de probabilidad multivariante.

Al comienzo se seleccionó una de las clases o lo que es lo mismo, un conjunto de parámetros de la red bayesiana. A continuación se ordenaron los nodos del grafo de acuerdo a un recorrido en anchura. Con ello se aseguró que se respetase la jerarquía de los nodos de manera que se muestrearon los valores de los nodos ancestros antes que los de sus descendientes. Este orden se denominó ancestral. A partir de este punto el muestreo consistió en generar valores de la distribución local de cada nodo siguiendo un proceso en cascada (Henrion, 1988).

Al muestrear a partir de una distribución de probabilidad gaussiana es posible que se generen valores negativos o muy extremos. Puesto que, por ejemplo, no tiene sentido definir la altura de un soma en números negativos, la distribución fue truncada. Se impusieron como límites del truncamiento para cada variable el valor mínimo y máximo encontrado en los somas reales, es decir, los utilizados en el proceso de clustering. Por lo tanto, para un nodo raíz se muestreó a partir de la distribución $\mathcal{N}(\mu, \sigma^2, a, b)$ donde μ es la media, σ^2 es la varianza, a es el valor mínimo de para los somas de la clase muestreada de acuerdo a un “hard clustering”, esto es, un clustering en el que cada observación se asigna al cluster con mayor probabilidad de pertenencia y b es el valor máximo. Para los nodos descendientes, la distribución fue $\mathcal{N}(\mu + \beta_p \mathbf{X}_p, \sigma^2, a, b)$, donde la media fue una regresión lineal de los padres siendo β_p el vector de pesos de los padres y \mathbf{X}_p el vector de valores muestreado para los padres.

Repitiendo el proceso se obtuvo un conjunto de datos de somas virtuales para cada una de las 3 clases. Cada observación definió un conjunto de características con las que generar un soma tridimensional.

5.1.2. Visualización 3D

La representación visual de las características muestreadas se fundamentó en dos fases. La primera de ellas consistió en un proceso iterativo con el que se situaron los centroides de las curvas entre secciones para a continuación generar la curva a su alrededor. La segunda se basó en triangular los puntos entre pares de secciones

consecutivas para formar la superficie del soma resultando una malla 3D (Figura 5.1).

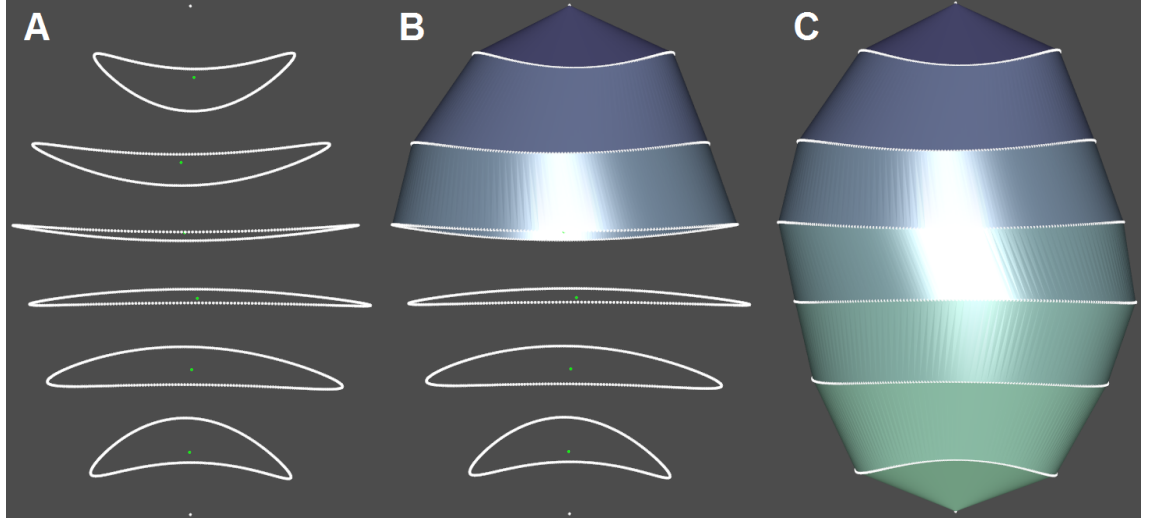


Figura 5.1: Proceso de visualización 3D. (A) Los puntos blancos representan las curvas entre secciones y los verdes sus centroides. En sus extremos el soma es en realidad un punto en lugar de una curva ya que cuando la colatitud es 90° o 270° todos los rayos trazados definen la misma línea recta que intersecta al soma en un único punto. (B) Se muestran las tres primeras secciones una vez se han triangulado los puntos de las curvas que imponen sus límites. (C) Malla triangular al finalizar el proceso.

5.1.2.1. Cálculo de las curvas entre secciones

El procedimiento comenzó situando el septentrión del soma en el origen de coordenadas (ver Figura 5.1). Tomando como referencia inicial ese punto se fueron calculando los centroides del resto de curvas siguiendo un proceso markoviano de orden 1 de manera que la posición del centroide siguiente solo dependiese del centroide anterior. La posición del centroide $i + 1$ a partir del centroide i se obtuvo operando con los descriptores α_{i+1} , ϕ_{i+1} y $|\vec{h}_{i+1}|$ que se relacionaron directamente con el vector \vec{h}_{i+1} , que une los centroides i e $i + 1$. α_1 y ϕ_1 fueron el resultado de medir ángulos sobre el vector $\vec{h}_{i+1} \times \vec{Z}$ donde $\vec{Z} = (0, 0, 1)$ es el vector unitario sobre el eje de cotas.

Para definir un vector se precisa de su longitud y orientación. Puesto que la longitud del vector \vec{h}_{i+1} es una de las variables simuladas, únicamente se precisó el cálculo de su orientación. Para recuperar la orientación de h_{i+1} en primer lugar se

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

calculó la magnitud del vector $|\vec{h}_{i+1} \times \vec{Z}| = r$ a partir de la relación del producto cruzado $\vec{h}_{i+1} \times \vec{Z}$ con α_{i+1} tal que $r = |\vec{h}_{i+1} \times \vec{Z}|^2 = |\vec{h}_{i+1}|^2 |\vec{Z}|^2 (1 - \cos \alpha_{i+1})^2 = |\vec{h}_{i+1}| |\vec{Z}| |\sin \alpha_{i+1}|$. La posición en el espacio del centroide $i+1$ se obtuvo convirtiendo las coordenadas esféricas ϕ_{i+1} y $\theta = \frac{\pi}{4}$ que determinaban la orientación del vector a coordenadas cartesianas. θ fue constante porque \vec{Z} también lo es. En disposición de la longitud y orientación del vector perpendicular la transformación fue como sigue:

$$x = r \sin \theta \cos \phi; \quad y = r \sin \theta \sin \phi; \quad z = r \cos \theta.$$

A partir de $\vec{h}_{i+1} \times \vec{Z}$ el cálculo de \vec{h}_{i+1} fue como sigue:

1. Se calculó el determinante que originó el vector $\vec{h}_{i+1} \times \vec{Z}$.

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \vec{h}_{(i+1)X} & \vec{h}_{(i+1)Y} & \vec{h}_{(i+1)Z} \\ 0 & 0 & 1 \end{vmatrix} = \vec{h}_{(i+1)Y} \vec{i} - \vec{h}_{(i+1)X} \vec{j}$$

Como se puede observar, el vector perpendicular $\vec{h}_{i+1} \times \vec{Z}$ únicamente intercambia las posiciones X e Y del vector \vec{h}_{i+1} . Por lo tanto $\vec{h}_{(i+1)X} = [\vec{h}_{i+1} \times \vec{Z}]_Y$ y $\vec{h}_{(i+1)Y} = [\vec{h}_{i+1} \times \vec{Z}]_X$.

2. Para el cálculo de $\vec{h}_{(i+1)Z}$ se utilizó la definición de producto escalar

$$\vec{h}_{i+1} \cdot \vec{Z} = |\vec{h}_{i+1}| |\vec{Z}| \cos \alpha_{i+1}$$

. En consecuencia $\vec{h}_{(i+1)Z} = |\vec{h}_{i+1}| \cos \alpha_{i+1}$. Puesto que se conocían tanto α_{i+1} como $|\vec{h}_{i+1}|$ se recuperó \vec{h}_{i+1} .

El proceso se repitió para todos los vectores \vec{h}_{i+1} . Al situar los vectores uno tras otro en orden de acuerdo al proceso markoviano se fueron describiendo las posiciones que ocupaban los centroides de las curvas que componían al soma.

El centroide de las curvas determinó la posición del centroide de las elipses sobre el plano XY . Mediante los ejes mayor y menor de las elipses así como su ángulo ω_i se pudo delimitar la forma de la curva i sobre el plano XY (ver Figura 3.2). La longitud de los ejes se calculó a partir de los descriptores r_i^h , que relacionaban el eje mayor de la elipse con la altura de las secciones, y e_i que es la proporción entre los ejes de la elipse.

Con la información correspondiente a la longitud de los ejes y al ángulo de la elipse ω_i se muestrearon coordenadas cartesianas de la elipse mediante:

$$x = \vec{h}_{iX} + B_i^R \cdot \cos \omega_i \cdot \cos U - B_i^r \cdot \sin \omega_i \sin U,$$

$$y = \vec{h}_{iY} + B_i^R \cdot \cos \omega_i \cdot \cos U + B_i^r \cdot \sin \omega_i \sin U,$$

donde $U \sim \text{unif}(0, 2\pi)$.

Finalmente, la profundidad de la curva se calculó a partir de los puntos muestreados de la elipse y los coeficientes β_{ij} :

$$z = \vec{h}_{iZ} + x^2 \cdot \beta_{i1} + y^2 \cdot \beta_{i2}.$$

El resultado final de este proceso se muestra en la Figura 5.1A.

5.1.2.2. Generación y reparación de las caras de las secciones

Una malla tridimensional se define a partir de sus vértices y caras. En este caso las curvas representaron la morfología local del soma por lo que sus puntos identificaron los vértices del soma. Las aristas entre los vértices del soma fueron las que caras determinaron la superficie de la malla. Para generar las caras de cada una de las secciones se enumeraron los vértices de la curva superior desde 1 hasta N siendo 1 un vértice escogido aleatoriamente. Desde el vértice 1 se calculó la distancia a los puntos de la otra curva que delimitaba la sección escogiendo el más cercano y asignándole la primera posición para la curva inferior. A partir de ese vértice también se enumeró la curva inferior. A continuación se generaron dos triángulos, cuyos vértices fueron por un lado el vértice 1 de la curva superior y el 1 y N de la inferior mientras que el segundo quedó definido por el 1 de la superior y el 1 y 2 de la inferior. El proceso de triangulación se extendió para los vértices impares de la curva superior de manera que para el vértice superior j los triángulos fueron j de la superior con j y $j-1$ de la inferior y j de la superior con j y $j+1$ de la inferior lo que dio como resultado la Figura 5.3A. De acuerdo a la enumeración de las curvas, un proceso semejante al realizado para la curva superior se realizó con la curva inferior pero en este caso con los vértices pares cuya representación se muestra en la Figura 5.3B. Al combinar los triángulos de la curva superior con los de la inferior se obtuvo una sección triangulada como la presentada en la Figura 5.3C.

Los somas generados mediante este método presentaron una morfología aproximada del soma en la que se distinguía claramente cada una de las curvas. Con el fin de aproximar la visualización de estos somas a la de los somas originales se muestrearon puntos en la superficie del soma mediante MeshLab (Cignoni). A partir de esos puntos se aplicó el método para la reconstrucción de superficies (2.2.1) obteniendo de esta forma una representación más próxima a la de un soma real. El resultado final de los somas simulados se muestra en la Figura 5.4.

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

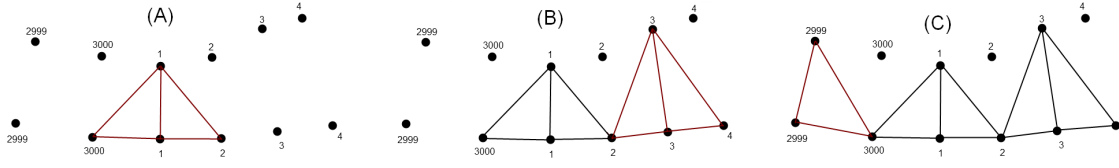


Figura 5.2: Triangulación de los puntos de las curvas. (A) Una vez se ha elegido un vértice de la curva superior aleatoriamente se generan dos triángulo con ese punto uniéndolo con el más cercano de la otra curva y los que se encuentran a ambos lados de este segundo punto. (B) Se triangulan el siguiente vértice impar. (C) Finaliza la triangulación para la curva superior.

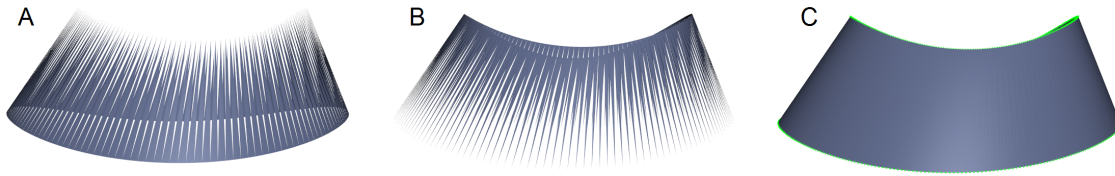


Figura 5.3: Generación de las caras. (A) Se triangulan los vértices de la curva superior. (B) Se triangulan los vértices de la curva inferior. (C) Combinación de (A) y (B).

5.2. Interpretación del modelo

5.2.1. Características segregadoras

5.2.1.1. Descriptores del Ray-tracer graph

Con el objetivo de identificar las características definitorias de cada cluster se realizó un conjunto de contrastes de hipótesis ANOVA. La selección de este método estadístico se vió motivada por la especificación de la red bayesiana dinámica espacial según la cual las variables seguían una distribución gaussiana. Por lo tanto, se compararon las poblaciones de cada clase contra el resto de clases variable por variable. Se asumió como hipótesis nula H_0 que los somas de distintas clases provenían de la misma población. Los descriptores para los que el p -valor fue inferior a 0,05 rechazaron la hipótesis nula. En esos casos se comparó el valor de la media para la clase evaluada respecto a las otras clases. En los casos en que fue superior se indicó mediante un signo positivo, en los casos en los que fue inferior con un signo negativo. Por lo tanto aquellas clases en las que se mostró un descriptor con el signo + se indicó que la clase destacaba por mostrar valores superiores al resto

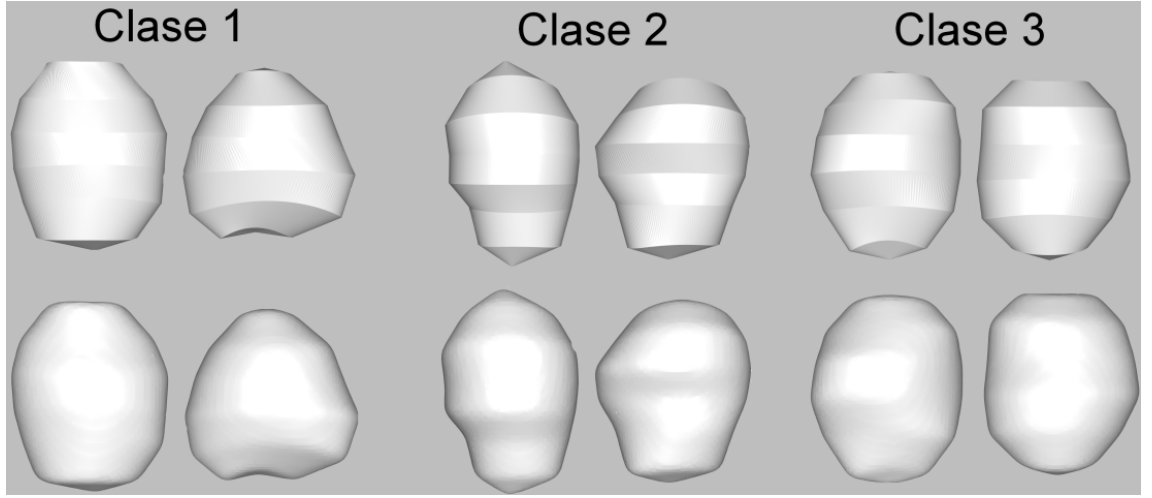


Figura 5.4: Ejemplos de somas simulados. En la representación se muestran ejemplos por pares de los distintos clusters. En la parte superior se muestra el resultado del proceso de generación de las caras entre las elipses que conforman las secciones mientras que inmediatamente debajo se presenta el final del proceso tras el muestreo y la aplicación del algoritmo de reconstrucción de superficies.

en ese aspecto mientras que el signo — señaló una diferencia destacable respecto al resto de clases pero en este caso por asumir valores más pequeños. Los resultados se muestran en la Tabla 5.1.

Por las características mostradas en la Tabla 5.1 se concluyó la siguiente caracterización.

- Clase 1: Se trata de somas achatados en su parte central y que crecen, visto desde el plano XY , siguiendo la dirección entre los $\frac{\pi}{2}$ y π radianes.
- Clase 2: Somas cuya mitad superior es especialmente alta y que crecen en sentido opuesto a la Clase 1, es decir, desde la perspectiva del plano XY entre los $\frac{3\pi}{2}$ y los 2π radianes.
- Clase 3: Estos somas se diferencian porque en su parte intermedia la dirección de crecimiento es una mezcla entre las clases 1 y 2 y porque las elipses de su parte superior se orientan en una dirección distinta a la de las otras clases.

Por lo tanto se puede concluir que las diferencias entre clases se centran esencialmente en los ángulos que forman las secciones y las elipses. De nuevo se motiva la aplicación de una distribución angular específica para tratar este tipo de variables que parecen tener un impacto importante en la discriminación de clases.

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

	Cluster 1	Cluster 2	Cluster 3
$ h_3 $	·	+	·
r_3^h	+	·	·
r_4^h	+	·	·
r_5^h	+	·	·
e_4	-	·	·
α_6	+	·	·
ϕ_2	+	-	·
ϕ_3	+	-	·
ϕ_5	-	+	-
ϕ_6	-	+	+
ϕ_7	+	-	·
ω_2	·	-	·
ω_3	·	·	-
ω_4	·	·	-
β_{13}	-	·	·
β_{23}	+	·	·

Tabla 5.1: Características discriminatorias entre clases. Se muestran las características con un p -valor inferior a 0,05 tras hacer el contraste de hipótesis ANOVA. El signo + identifica los casos en los que la media de la clase asume valores mayores para un descriptor que el resto de clases. El signo - indica que su valor es inferior al de las otras 2 clases. · significa que no rebasó el umbral para rechazar la hipótesis.

5.2.1.2. Entre áreas

Del mismo modo que se utilizó un conjunto de contrastes de hipótesis para estudiar las características definitorias de cada clase, el análisis fue extendido con el fin de averiguar si existen diferencias entre los somas extraídos de distintas áreas. En esta ocasión en lugar de aplicar un contraste de hipótesis ANOVA se utilizó el test de Kruskal-Wallis debido a que se desconoce la distribución que siguen los datos para cada área. De nuevo se compararon las poblaciones de cada área contra el resto de áreas variable por variable. Asimismo se asumió como hipótesis nula H_0 que los somas de distintas áreas pertenecían a la misma población. También se impuso como umbral de rechazo de H_0 un p -valor inferior a 0,05. Se compararon las medianas de los descriptores que se encontraron por debajo del umbral con la mediana del resto de las áreas para esa variable. Cuando la hipótesis fue rechazada para un descriptor

y el valor de la mediana de el área analizada mostró valores superiores al resto se indicó mediante el signo +, en el caso contrario se señaló mediante el signo -. Los resultados se muestran en la Tabla 5.2

	Cingular	Frontal	Temporal
r_4^h	·	+	·
e_2	+	-	·
e_3	+	-	·
e_4	·	-	·
e_5	·	-	·
e_6	·	-	·
α_4	+	·	-
α_5	·	+	·
ω_5	+	·	-
β_{12}	+	-	·
β_{13}	+	-	·
β_{14}	-	+	·
β_{15}	·	+	·
β_{16}	·	+	·
β_{22}	·	+	·
β_{23}	-	+	·
β_{24}	·	-	·
β_{25}	·	-	·
β_{26}	·	-	·

Tabla 5.2: Características discriminatorias entre áreas. Se muestran las características con un p -valor inferior a 0,05 tras hacer el test de hipótesis de Kruskal-Wallis. El signo + identifica los casos en los que la mediana del área analizada toma valores mayores para un descriptor que en el resto de áreas. El signo - indica que la mediana del área es inferior al de las otras 2 áreas. · representa aquellos casos en los que no se rebasó el umbral para rechazar la hipótesis nula.

A partir de los resultados mostrados en la Tabla 5.2 se identificaron propiedades que distinguieron a los somas de distintas áreas.

- Cingular: Las curvas superiores presentaron formas más redondeadas y menos elípticas que en el resto de áreas. Además, las curvas superiores presentan gran curvatura cuando se extienden mucho sobre el eje X y menor curvatura cuando se extienden poco sobre el eje X.

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

- Frontal: En su parte intermedia los somas son achatados y presentan curvas muy planas cuando el soma se extienden poco sobre el eje X y mucho sobre el eje Y y curvas con mucha pendiente cuando se extienden mucho sobre el eje X y poco sobre el eje Y .
- Temporal: Estos somas únicamente destacan porque en su mitad se produce un cambio de dirección en el crecimiento del soma y en que la dirección de una de sus elipses en la mitad sigue una dirección opuesta a la de la área cingular. Los resultados muestran que el área temporal es la que presenta mayor dificultad para ser distinguida de las otras dos. Esto puede deberse al reducido número de somas por lo que en futuros estudios en los que se disponga de mas somas debería de ser una propiedad a estudiar en profundidad.

5.2.1.3. Variables emergentes

En el Capítulo 3 se introdujeron un conjunto de características extraídas del estado del arte que fueron denominadas variables emergentes. Con el análisis de las diferencias entre grupos de somas obtenidas a partir del clustering de acuerdo a estas propiedades se persiguió una caracterización más completa de los grupos de somas. Para ello se aplicó un test Kruskal-Wallis sobre las variables emergentes y los grupos de somas. De nuevo se impuso como umbral de significatividad el p -valor 0,05. Al finalizar el proceso no se hallaron diferencias significativas entre las clases de acuerdo a las características globales.

El estudio se extendió aplicando el test Kruskal-Wallis sobre las variables emergentes de acuerdo a las áreas de las que se extrajeron los somas e imponiendo nuevamente un umbral de rechazo de hipótesis para el p -valor de 0,05. En esta ocasión sí que se hallaron diferencias entre cada una de las áreas como se muestra en la Tabla 5.3. Las conclusiones son:

- Cingular: Los somas de este área son más pequeños que los de las otras dos áreas debido a que muestran menor distancia entre los vértices y el centroide.
- Frontal: Estos somas se encuentran más achatados que los de las otras dos áreas aproximándose más a un elipsoide de acuerdo a las variables elipticidad y cociente entre los ejes.
- Temporal: La morfología de los somas de este área es más cercana a una esfera que el área frontal dado que su convexidad y solidez es mayor.

	Cingular	Frontal	Temporal
Esfericidad	.	-	.
Elongación	.	+	.
Distancia media sobre X	-	+	+
Distancia media sobre Y	-	+	+
Distancia media sobre Z	-	+	+
σ de la distancia sobre X	-	+	+
σ de la distancia sobre Y	-	+	+
σ de la distancia sobre Z	-	+	+
Convexidad	.	.	+
Solidez	.	-	+
Número vértices	+	-	-
Compacidad	.	+	.
Elípticidad	.	+	.
Anchura	.	-	.
Ratio ejes	.	+	.
log Ratio ejes	.	+	.

Tabla 5.3: Características discriminatorias globales entre áreas. Se muestran las características con un p -valor inferior a 0,05 tras hacer el test de hipótesis de Kruskal-Wallis. El signo + identifica los casos en los que la mediana del área analizada toma valores mayores para un descriptor que el resto de áreas. El signo - indica que la mediana del área es inferior al de las otras 2 áreas. . indica que no rebasó el umbral para rechazar la hipótesis.

5.2.2. Clasificación del soma de acuerdo a la similitud morfológica

En el Capítulo 3 se extendió la definición de esfericidad para cuerpos platónicos. La esfericidad, como se indicó, consiste en comparar el área de la superficie de un objeto tridimensional con el área de la superficie de una esfera del mismo volumen. El mismo cálculo se realizó para la pirámide, el cubo, el octaedro, el dodecaedro y el icosaedro.

Una relación exacta entre el objeto estudiado y el cuerpo platónico se obtiene cuando $\frac{A_o}{A_c} = 1$ donde A_o es el área del objeto y A_c es el área del cuerpo platónico, es decir, en caso de cumplirse la igualdad el objeto y el poliedro con el que se compara comparten el mismo volumen y el mismo área de la superficie. A partir de esta

5. SIMULACIÓN DE SOMAS E INTERPRETACIÓN DEL MODELO

relación es posible definir un método de clasificación de manera que la morfología del soma se identifique con el cuerpo platónico con el que mantenga la relación más próxima a 1. La asignación de cada uno de los somas al cuerpo platónico más similar se muestra en la Figura 5.5.

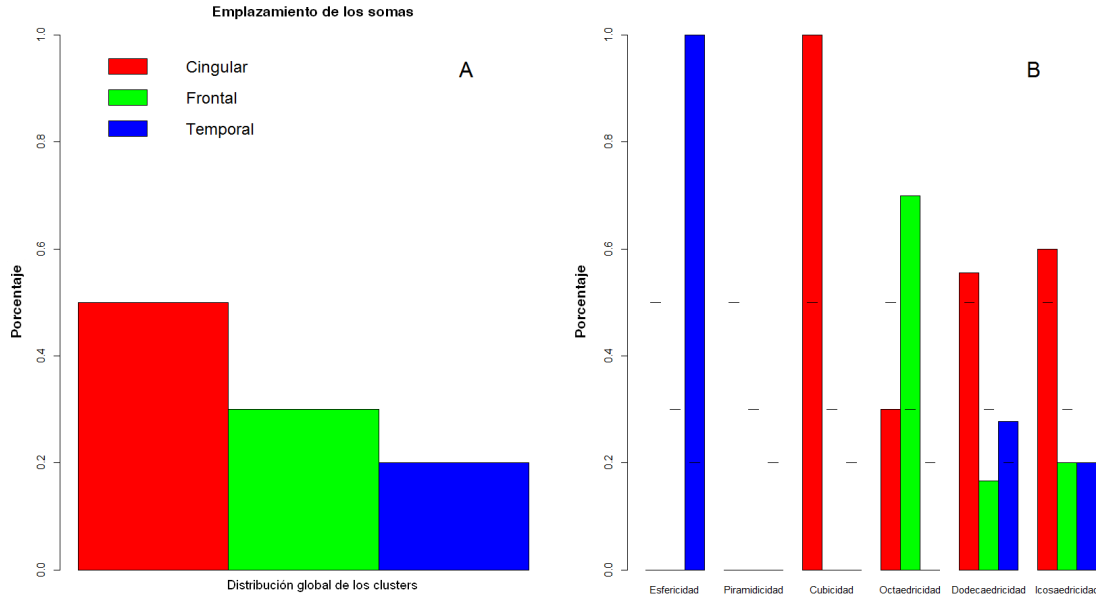


Figura 5.5: Clasificación de somas de acuerdo al área cerebral y la similitud morfológica. Diagrama de barras que muestra la distribución de los somas una vez han sido asignados al cuerpo platónico más parecido de acuerdo a las definiciones de piramidicidad, octaedricidad, etc. (A) Se muestra la distribución de probabilidad para cada uno de los somas a partir del área cerebral que ocupan. (B) Se presentan las distribuciones de los somas en las áreas cerebrales de acuerdo al cuerpo platónico más similar. Las líneas horizontales en (B) denotan las alturas de las barras mostradas en (A).

El resultado más destacado de la representación es la inexistencia de somas en la clase piramidicidad dado que los somas estudiados reciben el nombre de somas piramidales. También es reseñable que la morfología de la mayoría de los somas se aproxime más a los cuerpos con más caras, esto es, al octaedro, dodecaedro e icosaedro que a los cuerpos más simples. Esto puede deberse fundamentalmente a que un poliedro con más caras puede ajustarse mejor a formas más complejas. Una distribución que contrasta con las demás es la presentada en el área frontal donde la mayoría de los somas se aproximan a octaedros al contrario que en las demás áreas. En cualquier caso se debe de considerar que este análisis fue realizado con un número reducido de somas para las áreas temporal y frontal y que sería necesario disponer de más instancias de ambas clases para poder validar estos resultados.

Parte II

Espinas

6

Obtención, preprocesamiento y caracterización de las espinas

La segunda parte extendió el estudio de la morfología neuronal con el análisis de las espinas dendríticas. Al igual que para los somas, el desarrollo del trabajo siguió la metodología propuesta en Fayyad et al. (1996). A partir de ese procedimiento, las etapas de obtención, preprocesamiento y caracterización son fases diferenciadas. Sin embargo han sido agrupadas debido a que las dos primeras fueron realizadas en Baguear (2011) y por tanto son expuestas brevemente. La caracterización fue realizada como continuación de esos resultados.

6.1. Introducción

La relación entre la forma y la función de las espinas ha sido estudiada ampliamente desde el análisis cuantitativo demostrando fuertes correlaciones entre las variables morfológicas de las espinas y la estructura sináptica. Por ejemplo, el área de la densidad post-sináptica (PSD) está correlada con el número de vesículas pre-sinápticas, el número de receptores postsinápticos y la piscina de transmisores listos para ser liberados. También se conoce que la longitud del cuello de la espina es proporcional a la medida de aislamiento bioquímico y eléctrico de la espina respecto a su dendrita padre (Nusser et al., 2001; Yuste y Denk, 1995; Yuste et al., 2000; ?). Los experimentos han demostrado que las espinas grandes pueden generar corrientes sinápticas más grandes que las espinas pequeñas (Matsuzaki et al., 2004). Además, las espinas dendríticas son estructuras dinámicas con fluctuaciones en su volumen y que parecen tener importantes implicaciones en la memoria y en la cognición (Bon-

6. OBTENCIÓN, PREPROCESAMIENTO Y CARACTERIZACIÓN DE LAS ESPINAS

hoeffer y Yuste, 2002; Dunaevsky et al., 1999; Kasai et al., 2010; Matus, 2000). Por lo tanto, la morfología de las espinas dendríticas parecen tener una función crítica desde el punto de vista funcional.

Las espinas dendríticas muestran una gran variedad de morfologías, especialmente en el cortex humano (Benavides-Piccione et al., 2012) de las que se han propuesto diferentes clasificaciones de acuerdo a sus características. La de uso más habitual es la propuesta por Peters y Kaiserman-Abramof que categoriza las espinas en cuatro tipos esenciales: *Thin*, *Mushroom*, *Filopodium* y *Stubby* (Peters y Kaiserman-Abramof, 1970) (ver Figura 1.3). Sin embargo, se ha discutido si la gran diversidad de tamaños de espinas refleja en realidad un continuo de morfologías más que la existencia de clases discretas (Herrera-Arellano et al., 2007). En un trabajo reciente (Benavides-Piccione et al., 2012), se detalló la morfología de las espinas dendríticas del cortex cerebral humano. Una base de datos extensa y cuantitativa de espinas 3D completamente reconstruidas de las neuronas piramidales corticales reveló cambios selectivos en los parámetros dendríticos y de las espinas relacionados con el envejecimiento. También se hallaron correlaciones entre la morfología de las dendritas y de las espinas. Concretamente, con la edad se perdieron espinas cortas y pequeñas de las dendritas basales y espinas largas de las dendritas apicales independientemente de la distancia al soma. Asimismo, se ajustó una distribución teórica de probabilidad para los volúmenes y las longitudes de espinas dendríticas.

6.2. Extracción de los datos

Para el análisis se utilizó un conjunto de 7342 espinas dendríticas 3D reconstruidas individualmente a partir de las neuronas piramidales de la capa III del cortex cingular de dos varones humanos (con 40 y 85 años de edad). El tejido (suministrado por Dr. I. Ferrer, Instituto de Neuropatología Servicio de Anatomía Patológica, IDIBELL-Hospital Universitario de Bellvitge, Barcelona, España) fue obtenido durante la autopsia (2-3h post-mortem). El cerebro fue inmediatamente inmerso en frío 4% paraformaldeído en un tampón de fosfato de 0.1M, pH 7.4 (PB) y seccionado en porciones coronales de 1.4cm de anchura. A continuación, se transfirieron pequeños bloques del cortex (15mm × 10mm × 10 mm) a una segunda solución de 4% paraformaldeído en PB durante 24h a 4°C. Después de su fijación, se realizaron secciones de las circunvoluciones de la cingular anterior (área de Brodmann 24) con el vibratome. Después, las neuronas piramidales fueron intracelularmente inyectadas

con Lucifer Yellow (LY) y teñidas inmunofluorescentemente utilizando un anticuerpo contra el LY (ver Benavides-Piccione et al. (2012) para más detalles metodológicos). Posteriormente las dendritas apicales y basales fueron escaneadas con un microscopio confocal y completamente reconstruidas en tres dimensiones utilizando la metodología previamente descrita (Benavides-Piccione et al., 2012). Se obtuvieron imágenes (52 ± 17 imágenes, z-step de $0.28 \mu\text{m}$) con alta magnificación ($\times 63$ glycerol) para capturar la profundidad, longitud y anchura de las espinas en las dendritas apicales y basales (Figura 6.1). Las imágenes extraídas mediante el microscopio fueron analizadas mediante Imaris-Bitplane AG. Por medio de este software se obtuvieron representaciones tridimensionales de las espinas. Como se señala en Baguear (2011), la intensidad de la fluorescencia medida por el microscopio confocal varía a lo largo del tejido y aquellas partes pequeñas o finas que emiten poca luz en ocasiones no se distinguen del ruido. Esto motiva que sea necesario un procesamiento manual que implica seleccionar la intensidad de la imagen para cada espina y marcar el punto de inserción de la espina en el cuerpo dendrítico. Como resultado de esta tarea semiautomática se obtuvo un fichero VRML que representó la superficie de la espina dendrítica.

La distribución del conjunto de datos se presenta en la tabla 6.1.

Dendrita/Edad	C40	C85	
api	1993 (27,15 %)	1100 (14,98 %)	3093 (42,13 %)
bas	2550 (34,73 %)	1699 (23,14 %)	4249 (57,87 %)
	4543 (61,88 %)	2799 (38,12 %)	7342 (100 %)

Tabla 6.1: Distribución de las espinas: Distribución de las espinas del dataset de acuerdo a la edad del individuo del que se extrajeron dónde se representa un individuo de 40 años (C40) y otro de 85 años (C85) y el compartimento dendrítico ya sea apical (api) o basal (bas).

6.2.1. Reparación de las espinas

En Baguear (2011) se aplicó un procedimiento de reparación que trató de preservar la segmentación semiautomática con el objetivo de recuperar las espinas dañadas durante el proceso de fluorescencia. El proceso consistió en primer lugar en transformar las espinas que estaban representadas mediante mallas tridimensionales en un conjunto de voxels. Este proceso denominado voxelización utilizó el algoritmo presentado en Patil y Ravi (2005).

6. OBTENCIÓN, PREPROCESAMIENTO Y CARACTERIZACIÓN DE LAS ESPINAS

La segunda parte del proceso de reconstrucción se basó en un algoritmo evolutivo para estructuras 3D. Básicamente, el algoritmo evoluciona la espina rota siguiendo la intensidad de la pila de imágenes hasta que se alcanza un único volumen. El algoritmo que se aplicó es una mejora del algoritmo de la serpiente 2D presentado en Álvarez et al. (2010). La explicación de esta adaptación y el algoritmo completo se encuentran descritos en Baguear (2011).

En este caso el proceso de segmentación fue manual dado que las espinas son fácilmente identificables a partir de su cuello, es decir, el punto que las une a la dendrita.

6.3. Caracterización basada en Reeb graph

Al igual que con los somas se aplicó Reeb graph para poder obtener información acerca de las propiedades morfológicas globales y locales. Con ello se persiguió la construcción de una estructura del esqueleto del modelo tridimensional en forma de grafo que capturara la topología global mientras que, simultáneamente, se almacenaban características locales referentes a cada uno de los nodos del grafo. Debido a que las espinas muestran formas más diversas que en el caso de los somas y que el método de reconstrucción utilizado no reduce el número de puntos de la malla, se aplicó un método inspirado en el concepto de Reeb graph multiresolucional (MRG) (Tangelder y Velkamp, 2008) y su correspondiente implementación (Hilaga et al., 2001).

Esta técnica construye un grafo de un modelo geométrico 3D para describir su topología. La idea básica de este enfoque es dividir una malla triangular en regiones basadas en el valor de una función μ . Una selección habitual para μ es la distancia geodésica, que puede ser entendida como un problema en el que, dados dos puntos s y p en la superficie de un poliedro de n vértices, se trata de encontrar el camino mínimo sobre la superficie que permita alcanzar p desde s (O'Rourke, 1999). En el caso de las espinas s es un punto que fue marcado manualmente por un experto para identificar la cabeza de cada espina y p es cada vértice de la representación geométrica 3D.

El procedimiento más extendido para abordar este problema denominado mínimo camino geodésico es el algoritmo de Dijkstra continuo (Mitchell et al., 1987). Esta técnica consiste en propagar una señal desde el origen s hacia el resto de la superficie. Cuando la señal alcanza por primera vez un punto p se le asigna una

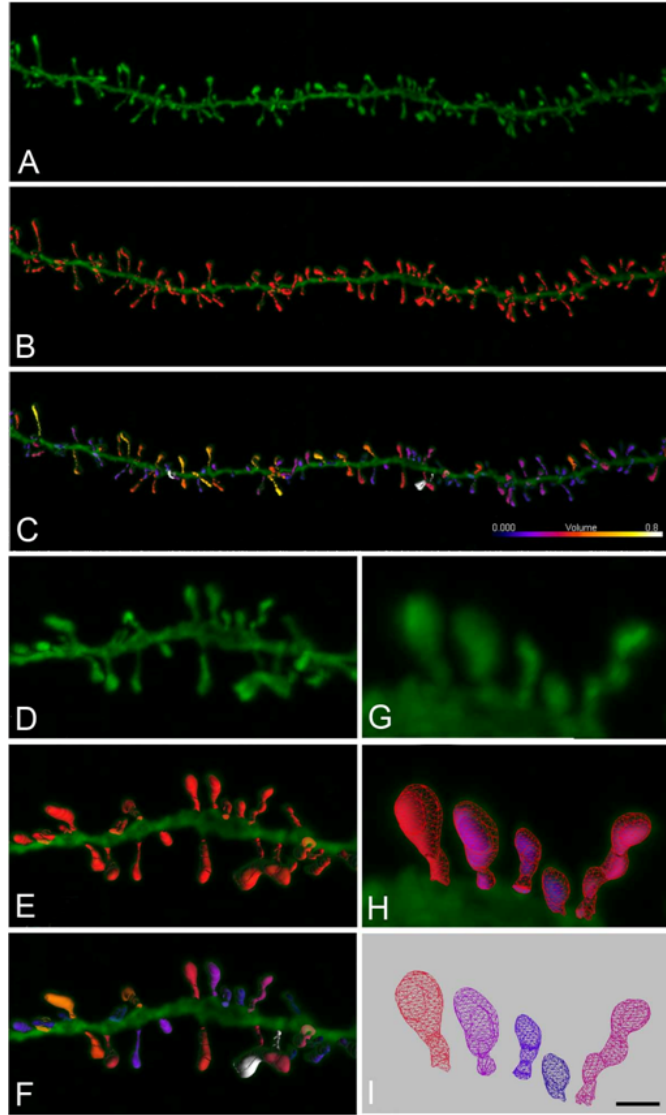


Figura 6.1: (A) Imagen de la proyección Z del microscopio confocal que muestra una dendrita basal horizontalmente proyectada extraída de una neurona piramidal del cortex cingular perteneciente al sujeto de 40 años de edad de la capa intracelular III tras ser inyectada. (B) Reconstrucción tridimensional de la morfología completa de cada espina dendrítica mostrada en (A). (C) Estimación del volumen de las espinas mostradas en (B) codificada por colores (azul-blanco: $0,0 - 0,8 \mu\text{m}^3$). (D-I) Alta magnificación de las imágenes del segmento dendrítico mostrado en A-C para ilustrar la malla triangular tridimensional (I) obtenida para cada espina individual.

etiqueta permanente con el tiempo $d(x)$ en el que recibió la señal. Dado que a nivel computacional el número de puntos que componen la malla es discreto, el

6. OBTENCIÓN, PREPROCESAMIENTO Y CARACTERIZACIÓN DE LAS ESPINAS

tiempo $d(x)$ se sustituye por la distancia del camino mínimo. Para el desarrollo de este trabajo se utilizó una implementación en Matlab que se puede encontrar en Exact geodesic for triangular meshes.

Formalmente, dada la representación de los límites de una espina $B \in \mathbb{R}^3$, se define la función continua $\mu : B \rightarrow \mathbb{R}$ como la longitud del camino mínimo desde la cabeza de la espina a un vértice arbitrario de la misma. El dominio de $\mu(\cdot)$ se divide en K intervalos r_i tales que $r_0 = [0, \frac{1}{K}]$, $r_1 = [\frac{1}{K}, \frac{2}{K}]$, ..., $r_{K-1} = [\frac{K-1}{K}, 1]$. Esto se traduce en que todos los vértices en la malla triangular se sitúan en una sección concreta $S_i, i = 1, \dots, K$, dependiendo de la función de evaluación. Esto puede provocar que varios triángulos queden comprendidos entre los límites de dos intervalos. Para evitar este problema y lograr cortes suaves entre secciones, esos triángulos son subdivididos añadiendo nuevos vértices dentro de cada arco afectado (Figura 6.2). La posición en la que se inserta un nuevo vértice se calcula a partir de la interpolación de la posición de los dos vértices que definen el eje afectado en la misma proporción que su distancia geodésica.

Como resultado, se obtiene un conjunto de regiones R_{ij} donde cada región es un conjunto de triángulos en una sección aislada de los otros conjuntos de la misma sección de manera que j indica la j -ésima región dentro de la i -ésima sección. Cada región representa un nodo del grafo el cual es una estructura de datos que contiene un conjunto de características topológicas locales. Los arcos del MRG se añaden entre las regiones originalmente conectadas. En el caso de las espinas, el MRG da lugar a una polilínea debido a su forma tubular. La Figura 6.3 muestra un ejemplo grafico para dos espinas dendríticas.

6.4. Descriptores de las espinas

Tras generar el MRG de cada una de las espinas se dispone para cada una de ellas de una superficie continua aproximada por una secuencia de $K = 8$ secciones $S_i, i = 1, \dots, 8$. Para cada sección se encuentran curvas definiendo los extremos superior T_i e inferior B_i . Al aproximar cada curva como una elipse, se asume que el eje mayor se define como T_i^R, B_i^R y el menor como T_i^r, B_i^r . Se debe de considerar en este proceso que la superficie requiere ser continua por lo que se imponen una serie de restricciones sobre secciones adyacentes: $\forall i, 1 < i < K + 1, B_i^R = T_{i-1}^R, B_i^r = T_{i-1}^r$. Para caracterizar a la espina se consideran las siguientes variables (Figura 6.4):

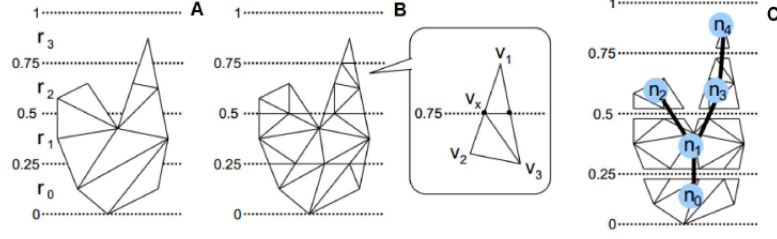


Figura 6.2: Modelo geométrico particionado y construcción del Reeb graph: (A) La malla triangular inicial es dividida en intervalos definiendo las secciones. (B) Los triángulos que se encuentran sobre los límites entre secciones son subdivididos añadiendo nuevos vértices (como v_x) dentro de cada arco afectado. (C) Cada sección S_i es dividida en un número arbitrario de regiones $R_{ij}, j = 1, \dots, K_i$, dependiendo de los componentes conectados de la partición. El MRG es construido mediante la representación de una estructura de datos que contiene un conjunto de características topológicas locales de la región R_{ij} como un nodo del grafo para esa región, y añadiendo arcos entre aquellos pares de regiones originalmente conectadas que pertenecen a secciones adyacentes.

- $|\vec{h}_i|, i = 1, \dots, K$: La altura relativa de cada sección. Es la longitud del vector \vec{h}_i definido por los centroides de dos elipses consecutivas.
- $a_i, i = 1, \dots, K$: El área de cada sección.
- $r_i^h, i = 1, \dots, K - 1$: El cociente $\frac{B_{i+1}^R}{|h_i| + |h_{i+1}|}$ donde B_{i+1}^R es el eje mayor de la elipse separando las secciones i -ésima y $(i + 1)$ -ésima y $|h_i|$ es la altura relativa de la sección i . Valores pequeños para r^h indican secciones elongadas mientras que valores elevados para r^h implican secciones achatadas.
- $e_i, i = 1, \dots, K - 1$: El ratio entre los ejes de la elipse $B_{i+1}^r : B_{i+1}^R$ es calculado a partir de $\frac{B_{i+1}^r}{B_{i+1}^R}$ donde B_{i+1}^r es el eje menor de la elipse separando las secciones i -ésima e $(i + 1)$ -ésima y el eje mayor B_{i+1}^R de la misma elipse. Cuando el ratio entre ejes de la elipse es igual a 1 la elipse es, de hecho, un círculo. Según se reduce el valor de e la elipse se va aplastando.
- $\cos \alpha_i, i = 1, \dots, K - 1$: El coseno del ángulo entre dos vectores. El primero de los vectores es el vector que conecta los centroides de las elipses que separan las secciones $(i - 2)$ -ésima e $(i - 1)$ -ésima y las secciones $(i - 1)$ -ésima e i -ésima. El segundo vector queda definido por los centroides que conectan las elipses que separan las secciones $(i - 1)$ -ésima y la i -ésima y las secciones

6. OBTENCIÓN, PREPROCESAMIENTO Y CARACTERIZACIÓN DE LAS ESPINAS

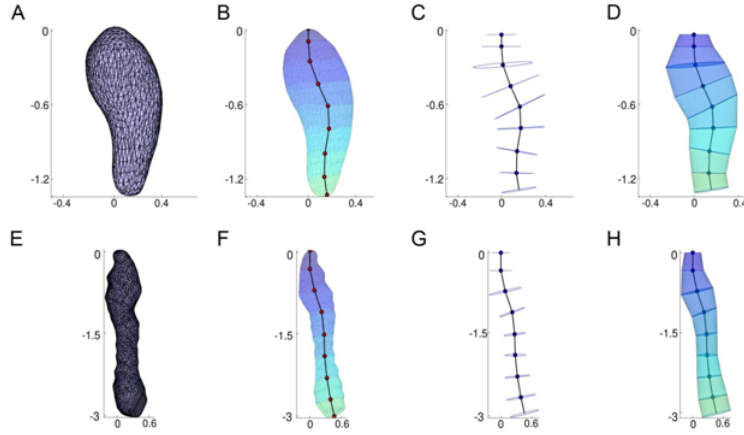


Figura 6.3: (A,E) Representación visual de dos ejemplos de espina dendrítica que muestran la malla triangular extraída manualmente después de su reconstrucción. (B,F) Grafos creados con MRG incluyendo la malla triangular original dividida por colores de acuerdo a un número determinado de secciones (en este caso ocho). Estas representaciones geométricas 3D han sido trasladadas de manera que los puntos que determinan las cabezas han sido colocados en el origen y las primeras secciones son alineadas con el eje Z. (C,G) Esqueleto y elipses a partir de las cuales se extraen los vectores de características. (D,H) Reconstrucción de las espinas de acuerdo a la envoltura convexa de cada sección.

i -ésima e $(i + 1)$ -ésima. El coseno es calculado a partir del producto escalar $\cos \alpha_i = \frac{h_i \cdot h_{i+1}}{|h_i||h_{i+1}|}$.

- $\theta_i, i = 2, \dots, K - 1$: El ángulo polar, también denominado colatitud en un sistema de coordenadas esféricas. El centroide de la elipse B_{i+1} se asume como centro de la esfera. La punta del vector $h_i \times h_{i+1}$ es el punto proyectado a partir del cual se calcularon el ángulo polar θ_i y el ángulo azimuth ϕ_i . El ángulo polar fue calculado a partir de:

$$\theta = \begin{cases} \arctan \left(\frac{\sqrt{x^2+y^2}}{z} \right) & z > 0 \\ \frac{\pi}{2} & z = 0 \\ \pi + \arctan \left(\frac{\sqrt{x^2+y^2}}{z} \right) & z < 0 \end{cases}$$

donde x, y y z son las coordenadas correspondientes a la punta del vector $h_i \times h_{i+1}$. Dado que h_1 se situó sobre el eje Z para orientar todas las espinas en la misma dirección, θ_1 es constante y por ese motivo no fue incluido en los pasos posteriores.

- $\phi_i, i, \dots, K-1$: El azimuth o ángulo azimutal, se obtiene tras la transformación al sistema de coordenadas esféricas. De nuevo se asumió que el centroide de la elipse B_{i+1} es el centro de la esfera y que la punta del vector $h_i \times h_{i+1}$ es el punto proyectado. El ángulo azimutal se calculó a partir de:

$$\phi = \begin{cases} \arctan\left(\frac{y}{x}\right) & x > 0, y > 0 \\ 2\pi + \arctan\left(\frac{y}{x}\right) & x > 0, y < 0 \\ \frac{\pi}{2} \operatorname{sgn}(y) & x = 0 \\ \pi + \arctan\left(\frac{y}{x}\right) & x < 0 \end{cases}$$

donde x y y son las coordenadas correspondientes a la punta del vector $h_i \times h_{i+1}$. La punta del vector es proyectada en el plano XY mediante análisis de componentes principales (PCA). A continuación, el vector se define a partir del centro de la esfera hacia el punto proyectado en XY. El ángulo azimuth se midió entre este nuevo vector y el eje X en el plano XY empezando a contabilizar desde el eje X en el sentido contrario a las agujas del reloj.

Como resultado, asumiendo $K = 8$, para cada una de las espinas se obtiene un total de 50 variables que permiten caracterizar la espina.

6. OBTENCIÓN, PREPROCESAMIENTO Y CARACTERIZACIÓN DE LAS ESPINAS

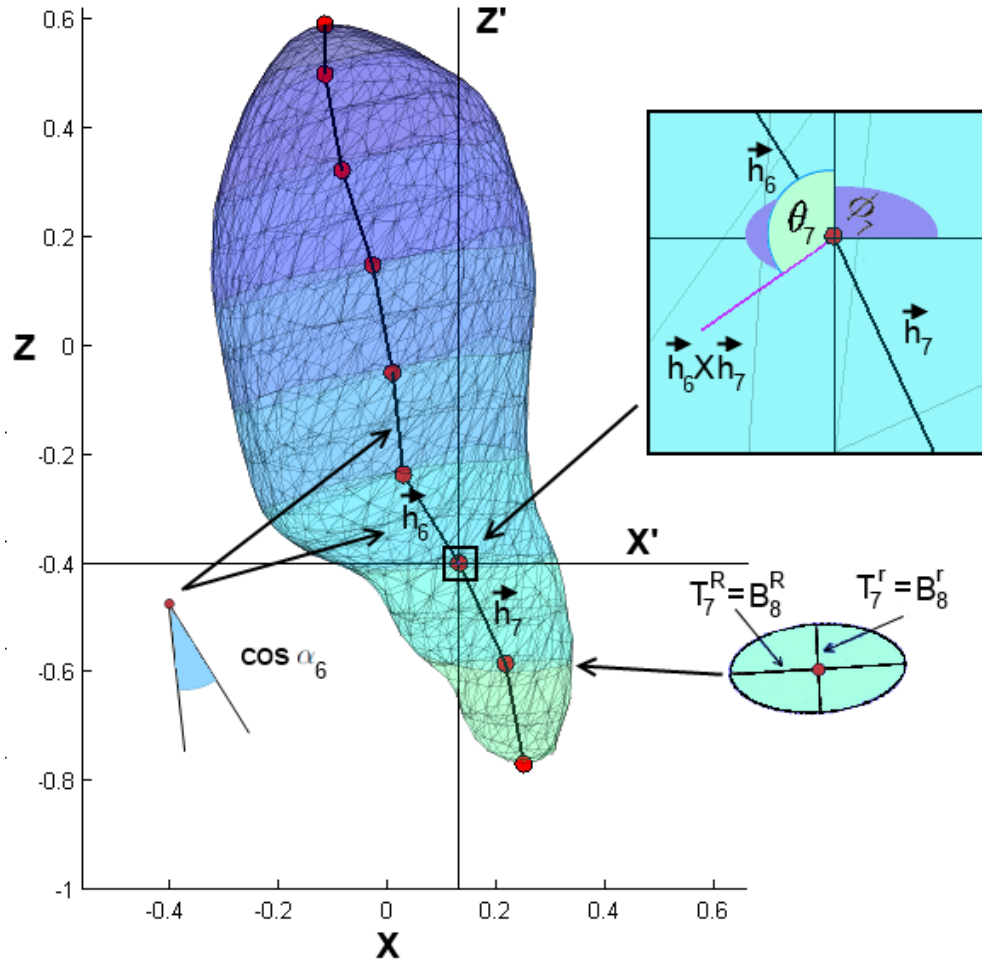


Figura 6.4: Descripción de las características de las espinas: Los centroides de las elipses se muestran como puntos rojos. Estos puntos están conectados por vectores cuya longitud es $|h_i|$. Una elipse se encuentra definida por sus centroide, el eje mayor ($T_{i-1}^R = B_i^R$) y el eje menor ($T_{i-1}^r = B_i^r$) el cual puede ser utilizado para calcular r_i^h y e_i . Los vectores \vec{h}_i y \vec{h}_{i+1} delimitan un plano cuyo ángulo es α . A partir de $\vec{h}_i \times \vec{h}_{i+1}$ se obtiene la punta del vector proyectado necesario para calcular θ y ϕ .

7

Modelización de las espinas

7.1. Definición del modelo

Llegado este punto, el problema que se pretende abordar consiste en definir y emplear un proceso para la obtención de un modelo matemático que permita identificar grupos de espinas. Por lo tanto el siguiente paso consiste en seleccionar un algoritmo de clustering y aplicarlo sobre los datos.

Para agrupar las distintas instancias que componen el conjunto de datos se seleccionó la mixtura de gaussianas como el procedimiento a seguir. Este enfoque basado en “soft clustering” se justifica debido a que, de acuerdo a Herrera-Arellano et al. (2007), las espinas definen un continuo de formas. La aplicación de un algoritmo de clustering cuyo resultado son probabilidades de pertenencia en oposición del “hard clustering” donde cada elemento pertenece exclusivamente a un grupo permite un análisis y entendimiento más profundo de los datos. Puesto que se desconoce la familia de la distribución subyacente a los datos, una elección común consiste en asumir que los datos siguen una distribución normal ya que, como señaló Pearson, es capaz de aproximar ciertas distribuciones y gracias a su simplicidad se puede utilizar como una primera aproximación. Una herramienta adecuada y que se ajusta al lenguaje de programación que se aplica a lo largo del trabajo es Mclust (Fraley y Raftery, 2002) ya que es un paquete perteneciente al R-project (R Development Core Team, 2008) desarrollado para el clustering basado en modelado y estimación de densidad de acuerdo a modelos de mixturas finitas de gaussianas.

7. MODELIZACIÓN DE LAS ESPINAS

7.1.1. Clustering probabilístico de espinas basado en mixtura de gaussianas

El clustering basado en modelos (McLachlan y Basford, 1988) es un enfoque de clustering probabilístico que asume que los datos de cada clase fueron generados por diferentes distribuciones de probabilidad. Cada distribución modela los datos de una clase distinta y el conjunto de todas las distribuciones se combina para dar lugar a un modelo finito de mixturas \mathbf{M} . Habitualmente los parámetros de un modelo finito de mixturas se estiman junto con el clustering solución de acuerdo a la estimación de máxima verosimilitud (EMV), esto es, se trata de encontrar el conjunto de parámetros $\boldsymbol{\theta}$ que maximiza la verosimilitud de los datos observados, es decir, $\arg \max_{\boldsymbol{\theta}} P_{\boldsymbol{\theta}}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ donde $\mathbf{x}_1, \dots, \mathbf{x}_N$ son los datos.

$$P(\mathbf{x}_1, \dots, \mathbf{x}_N | \mathbf{M}, \boldsymbol{\theta}) = \sum_{c=1}^C \pi_c \mathcal{N}(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$

donde \mathcal{N} denota una distribución normal con probabilidad a priori π_c , media $\boldsymbol{\mu}_c$ y cuya matriz de covarianzas es $\boldsymbol{\Sigma}_c$, C es el número total de Gaussianas y cada Gaussianiana se denota por c . El objetivo es obtener los valores para el conjunto de parámetros

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_c, \dots, \boldsymbol{\theta}_C)$$

$$\boldsymbol{\theta}_c = (\pi_c, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$

que maximiza la verosimilitud de los datos. Esto se puede conseguir de una forma aproximada, al igual que en el caso de los somas, utilizando el algoritmo esperanza-maximización (EM) (Dempster et al., 1977).

EM comienza inicializando los parámetros $\boldsymbol{\theta}$ y evaluando la log-verosimilitud de un modelo \mathbf{M} con los parámetros generados. A continuación, el paso de esperanza, se calcula la responsabilidad de la c -ésima Gaussianiana para la n -ésima observación, igual que en el caso del soma:

$$\gamma(c_n) = \frac{\pi_c \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}{\sum_{j=1}^C \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

El paso de maximización re-estima los parámetros utilizando las responsabilidades calculadas en el paso de esperanza. Así, los parámetros $\boldsymbol{\theta}$ son optimizados en

sus nuevas versiones de acuerdo a:

$$\begin{aligned}\boldsymbol{\mu}_c^{new} &= \frac{1}{N_c} \sum_{n=1}^N \gamma(c_n) \mathbf{x}_n \\ \boldsymbol{\Sigma}_c^{new} &= \frac{1}{N_c} \sum_{n=1}^N \gamma(c_n) (\mathbf{x}_n - \boldsymbol{\mu}_c^{new})(\mathbf{x}_n - \boldsymbol{\mu}_c^{new})^T \\ \pi_c^{new} &= \frac{N_c}{N}\end{aligned}$$

donde

$$N_c = \sum_{n=1}^N \gamma(c_n).$$

A continuación se evalúa la log-verosimilitud

$$\ln P(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\theta}^{new}) = \sum_{n=1}^N \ln \left[\sum_{c=1}^C \pi_c^{new} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_c^{new}, \boldsymbol{\Sigma}_c^{new}) \right].$$

El algoritmo EM itera hasta que o bien los parámetros o la log-verosimilitud convergen. Una vez ha finalizado EM y los parámetros para el modelo de mixturas han sido obtenidos, se devuelve la distribución de probabilidad de pertenecer a cada cluster para cada espina.

Para la selección del modelo se aplicó la puntuación BIC. Mclust calcula la puntuación BIC para cada uno de los modelos de acuerdo a Fraley y Raftery (2002) a partir de:

$$\begin{aligned}BIC_M &= 2 \ln P(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\theta}, \mathbf{M}) - v_M \ln(N) \\ v_M &= N \cdot C + \binom{N}{2} \cdot C + C\end{aligned}$$

donde $\boldsymbol{\theta}$ son los parámetros estimados mediante EM, \mathbf{M} es el modelo de mixtura de gaussianas, $\mathbf{x}_1, \dots, \mathbf{x}_N$ son los datos observados, N el tamaño de la muestra y v_M el número de parámetros.

7.2. Distribución de las espinas por cluster

Utilizando este enfoque el modelo para el que se obtuvo la mejor puntuación BIC fue en el que las espinas se desglosaban en siete grupos (Figura 7.1). En la representación se observan dos máximos, uno global de 7 grupos y otro local con 4 clases que coinciden con el número de espinas presentadas en la categorización de

7. MODELIZACIÓN DE LAS ESPINAS

Peters y Kaiserman-Abramof (1970). Ambos máximos fueron estudiados a lo largo de este trabajo tomando como referencia el modelo de siete grupos. El resultado del algoritmo asignó a cada espina una cierta probabilidad de pertenencia a cada grupo de manera que se obtuvo una distribución de probabilidad (p_1, \dots, p_7) para el clustering de 7 clusters y (p_1, \dots, p_4) para el clustering de 4 clusters, donde p_i es la probabilidad de pertenencia al grupo i y $p_1 + \dots + p_7 = 1, p_i \geq 0$. La probabilidad de pertenencia a un único grupo predominó para algunas espinas (con una probabilidad de pertenencia es superior a 0.99) mientras que para otras las probabilidades fueron más dispersas y su pertenencia no estuvo clara.

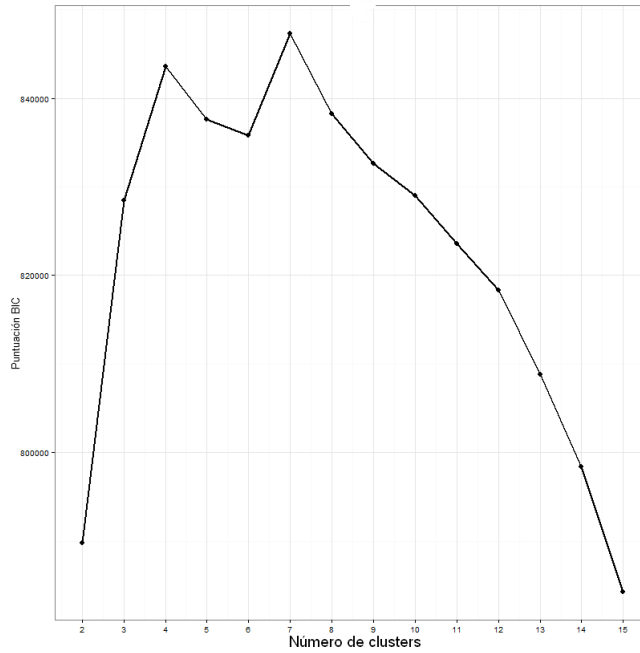


Figura 7.1: Gráfico que muestra los valores BIC obtenidos dependiendo del número de clases. Los resultados que se muestran se encuentran en el rango entre 2 y 15 clusters. Se puede observar que se alcanza un máximo global con 7 clusters y un máximo local con 4 clusters.

En Herrera-Arellano et al. (2007) se indica la existencia de una gran diversidad de morfologías en las espinas con lo que se observa más un continuo de espinas que la existencia de clases concretas. Con el fin de discutir si el conjunto de probabilidades del clustering para cada espina muestra un continuo de morfologías, en este trabajo se contabilizaron el número de espinas no clasificadas si la máxima probabilidad de todos los p_i , denotado como p^* , es menor que un umbral de probabilidad (Tabla 7.1).

7.2 Distribución de las espinas por cluster

Como se puede apreciar la mayoría de las espinas fueron asignadas al cluster más probable con una probabilidad superior al 0.99. Además ninguna de ellas fue clasificada con una probabilidad inferior al 0.41. En consecuencia existe un continuo entre las clases de las espinas ya que para algunas de ellas la probabilidad de pertenencia entre clases está equilibrada. Sin embargo, también es cierto que la mayor parte de las espinas quedan agrupadas bajo un cluster con una probabilidad muy elevada. Por lo tanto se puede concluir que a pesar de la existencia de un continuo de espinas sí que existen grupos de espinas.

Prob\Clust	Cluster 1 (1919)	Cluster 2 (772)	Cluster 3 (989)	Cluster 4 (731)	Cluster 5 (775)	Cluster 6 (1550)	Cluster 7 (606)
0.99	303	93	116	76	127	304	74
0.9	103	23	39	25	39	102	21
0.8	50	17	20	15	25	54	10
0.7	26	9	14	9	16	30	6
0.6	14	5	3	3	7	16	2
0.5	0	1	0	1	0	2	0
0.41	0	0	0	0	0	0	0

Tabla 7.1: Número de espinas no clasificadas que pertenecen a un cluster si la probabilidad máxima de pertenencia a un cluster p^* es inferior a un umbral. El número total de espinas para los cluster está entre paréntesis. La primera columna establece un umbral de probabilidad. Por ejemplo, 303 espinas de las 1919 que se encuentran en el cluster 1 obtuvieron una probabilidad p^* inferior a 0.99. Se debe de considerar que el número de espinas no clasificadas se reduce drásticamente según el umbral decrece. Además, de acuerdo a estos resultados, la mayoría de las espinas forman parte de un cluster con una probabilidad superior al 0.99.

Del mismo modo que se analizó la existencia de un continuo de espinas para 7 clusters, se analizó la distribución de probabilidad para 4 clusters imponiendo los mismos umbrales de probabilidad. El resultado obtenido se muestra en la Tabla 7.2.

7. MODELIZACIÓN DE LAS ESPINAS

Prob\Clust	Cluster 1 (2480)	Cluster 2 (1337)	Cluster 3 (1309)	Cluster 4 (2216)
0.99	276	154	151	335
0.9	92	65	75	125
0.8	51	43	37	67
0.7	37	21	26	34
0.6	14	10	13	18
0.5	0	1	1	0
0.4248	0	0	0	0

Tabla 7.2: Número de espinas no clasificadas que pertenecen a un cluster si la probabilidad máxima de pertenencia a un cluster p^* es inferior a un umbral. El número total de espinas para los cluster está entre paréntesis. Al igual que la Tabla 7.1 el número de espinas no clasificadas se reduce drásticamente según el umbral decrece. En este caso también la mayoría de las espinas forman parte de un cluster con una probabilidad superior al 0.99.

8

Interpretación del modelo

El algoritmo de clustering permite reconocer que elementos presentan características similares y agruparlos. Sin embargo, es labor del analista de datos identificar la relación de propiedades que caracterizan cada clase y las diferencias entre clases. Debido a la naturaleza de las características que se escogieron en la fase de preprocesamiento, la interpretación del modelo y en consecuencia la diferenciación entre grupos de espinas resultó una tarea compleja. A lo largo de este apartado se establecen los pasos que se siguieron para acometer esta cuestión.

8.1. Caracterización de las clases

Dado que tras la finalización del algoritmo de clustering fue posible asignar cada espina a la clase para la que se obtuvo un p^* mayor, las espinas fueron etiquetadas y se dispuso de más información con la que trabajar. Concretamente, el problema fue abordado desde el ámbito del aprendizaje supervisado con la intención de aplicar algún método que facilitase la comprensión de las clases. En este aspecto es común utilizar algoritmos basados en inducción de reglas, es decir, extraer reglas formales a partir de un conjunto de observaciones de forma que las instancias se asignen a la clase adscrita a la regla que cumplen.

8.1.1. Algoritmo RIPPER

Con el objetivo de arrojar luz acerca de la caracterización de cada clase, se trazó un proceso basado en la generación de reglas de clasificación de acuerdo a RIPPER (Cohen, 1995), algoritmo implementado en Weka (Hall et al.), para hallar las variables de cada clase que la diferencian del resto. RIPPER toma como punto de

8. INTERPRETACIÓN DEL MODELO

partida el algoritmo IREP (Fürnkranz, 1997) para generar reglas. RIPPER divide los datos en dos conjuntos, *conjunto de crecimiento* y *conjunto de poda*. El proceso algorítmico es dividido en crecimiento (generación de reglas) y poda (reducción del número de condiciones en una regla). Una regla crece añadiendo condiciones que maximicen el criterio de ganancia de información FOIL (Quinlan, 1990), esto es, se compara la información que aporta la regla antes y después de añadir una conjunción. Se calcula mediante

$$FOIL = t \cdot \left[\log_2 \left(\frac{p_1}{p_1 + n_1} \right) - \log_2 \left(\frac{p_0}{p_0 + n_0} \right) \right],$$

donde p_0 y p_1 son las instancias positivas para la regla sin y con la nueva conjunción respectivamente, n_0 y n_1 idem para las negativas y t es el número de instancias positivas cubiertas por ambas reglas.

Como resultado de la ejecución de RIPPER se obtiene una conjunción de condiciones de la forma $X = x$ para las variables nominales y $X \leq x$, o $X \geq x$ para las variables continuas. Después de que una regla sea generada comienza el proceso de poda que elimina las condiciones que maximicen $v(Regla, PodaPos, PodaNeg) = \frac{pos - neg}{pos + neg}$, donde pos (neg) es un número de ejemplos positivos (negativos) cubiertos por una regla. El proceso de poda es aplicado hasta que no hay más eliminaciones que mejoren v .

En este trabajo, mediante la selección del cluster más probable para cada espina, las espinas fueron adscritas a un único cluster. A continuación se generaron 7 conjuntos de datos, uno para cada cluster, en el que se utilizaron dos etiquetas. Una etiqueta designa una única clase mientras que la otra etiqueta se utiliza para identificar al resto de clases. De esta forma cada conjunto de datos denota un cluster distinto que se enfrenta al resto lo que permite que para cada cluster se obtengan reglas que la discriminen del resto.

Para lidiar con el desequilibrio en la cantidad de instancias de cada clase derivadas de la división de los datos (una clase contra el resto), se realizó un submuestreo aleatorio como paso de pre-procesamiento antes de la aplicación de RIPPER para tratar de evitar el sesgo. El submuestreo aleatorio es una técnica para ajustar la distribución de clases de manera que el número de observaciones de la clase más representada se reduce mediante la selección aleatoria de observaciones hasta que la proporción de clases se iguale. Con el objetivo de obtener las reglas más representativas y así facilitar la comprensión, se obligó al algoritmo a que el mínimo número de instancias clasificadas por regla fuera de al menos el 10 % del número total de espinas en ese cluster.

8.1 Caracterización de las clases

	$ h_2 $	$ h_4 $	$ h_5 $	$ h_6 $	$ h_7 $	a_1	a_4	a_5	r_2	r_3	r_4	r_5	r_6	r_7	e_3	e_5	e_6	α_4	α_5	α_6	α_7	ϕ_4	ϕ_5	ϕ_6
Cluster 1					X					X	X	X	X	X		X			X	X	X			
Cluster 2		X					X		X										X	X	X			
Cluster 3						X																		
Cluster 4												X			X									
Cluster 5				X																		X	X	X
Cluster 6	X		X	X		X		X			X		X				X	X						
Cluster 7			X			X	X						X									X	X	X

Tabla 8.1: Variables discriminatorias de acuerdo a RIPPER para cada clase con 7 clusters. Las variables que aparecen como mínimo una vez en el conjunto de reglas para una clase se denotan con una cruz. Al buscar una relación entre las variables y las espinas de la Figura 8.1 se observa que aquellas espinas que tienen cabeza están relacionadas con las variables definidas por los ángulos α y ϕ , esto es, la cabeza de la espina implica un cambio direccional entre dos secciones del MRG. También se observa que el área se utiliza para discriminar de acuerdo al tamaño así como la altura. Las variables r (achataamiento y elongación) y e (excentricidad) han resultado relevantes en la detección de cuellos dado que tienen valores extremadamente bajos en los clusters 4 y 7. En cuanto a la colatitud, no ha sido incluida en la tabla porque no se utiliza en ninguna regla; parece que no contribuye en la diferenciación entre grupos de espinas. Además, se puede ver la complejidad para describir algunos clusters ya que bastantes variables están implicadas en el proceso como es el caso de los clusters 1, 6 y 7 mientras que para otros cluster destaca la simplicidad al poder ser identificados con una o dos variables como los clusters 3 y 4.

8.1.2. Características segregadoras

Con el objetivo de identificar las características definitorias de cada cluster se generó un conjunto de reglas mediante RIPPER. Las variables características de cada cluster se muestran en las Tablas 8.1 y 8.2 para 7 clusters y en la Tabla 8.2 para 4 clusters. Para hacer posible su interpretación también se ha incluido en el Apéndice 9.2 un resumen de los valores que toman las características de las espinas para cada variable. Los conjuntos de reglas se detallan en el Apéndice 9.2.

A pesar de la aparente diversidad y complejidad de las reglas necesarias para disgregar espinas en clusters, su interpretación descubrió que cada prototipo de cluster está en realidad caracterizado por una o dos propiedades observables. En consecuencia, cada conjunto de reglas de cluster fueron descritas definiendo las espinas de la como se describe a continuación:

8. INTERPRETACIÓN DEL MODELO

	$ h_6 $	$ h_7 $	a_1	a_6	r_4	r_5	r_6	r_7	e_3	e_4	α_5	α_6	α_7
Clase 1		X	X	X	X		X	X		X	X	X	X
Clase 2						X	X		X	X			
Clase 3			X										
Clase 4	X	X				X		X					

Tabla 8.2: Variables discriminatorias de acuerdo al algoritmo RIPPER para cada clase con 4 clusters. Las variables que aparecen al menos una vez en el conjunto de reglas para una clase se denotan con una cruz.

8.1.2.1. Descripción de las espinas con 7 clases

- *Clase 1:* En la mitad superior (la parte más alejada de la dendrita) la espina es normal pero en su parte inferior (base de la espina) tiende a ser achatada y recta.
- *Clase 2:* Las espinas que tienen cabeza y en su parte media son muy largas pertenecen a este grupo.
- *Clase 3:* En este grupo se incluyen aquellas espinas que son extremadamente pequeñas en su parte superior.
- *Clase 4:* Aquellas espinas que tienen un cuello extremadamente pequeño en su parte intermedia están en este cluster.
- *Clase 5:* Las espinas cuya cabeza comienza aproximadamente en su parte intermedia. Además no son muy cortas en su parte inferior.
- *Clase 6:* Cuando las elipses que delimitan la parte intermedia de la espina son bastante redondeadas, la espina es bastante achatada en su parte intermedia, ligeramente estrecha en la parte inferior y bastante recta la espina pertenece a este cluster.
- *Clase 7:* Las espinas que muestran cabeza y tienen cuello en su parte intermedia se incluyen en este grupo.

Una vez las diferencias entre espinas han sido presentadas de acuerdo al resultado de RIPPER, con el objetivo de obtener una mejor comprensión del modelo, las diez espinas con mayor probabilidad de pertenencia para cada uno de los siete clusters se muestran en la Figura 8.1 con la finalidad de comprobar la correspondencia entre las reglas de RIPPER y la auténtica forma.

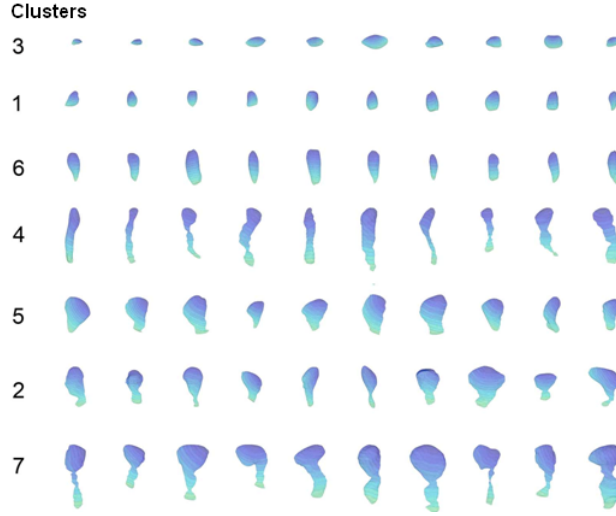


Figura 8.1: Las diez espinas dendríticas más representativas para cada cluster con $p^* \approx 1$ (probabilidad de pertenencia más alta), para los siete clusters. Como se puede comprobar, las características que indujo RIPPER coincidieron con las mostradas en la imagen lo que facilitó la diferenciación entre clusters.

8.1.2.2. Descripción de las espinas con 4 clases

Descripción de las espinas tras la interpretación de las reglas resultado de aplicar RIPPER sobre el modelo de 4 clusters:

- *Cluster 1*: Estas espinas son cortas y grandes en su base y no tienen ni cabeza ni cuello.
- *Cluster 2*: En este grupo se incluyen aquellas espinas que tienen un cuello estrecho y alargado en su parte intermedia.
- *Cluster 3*: Espinas extremadamente pequeñas en su parte más alta.
- *Cluster 4*: Espinas que en su base muestran cuello y en su parte intermedia no son alargadas.

Al igual que para el modelo de 7 categorías se muestran Figura 8.2 las diez espinas con probabilidad de pertenencia más alta para cada cluster.

8. INTERPRETACIÓN DEL MODELO

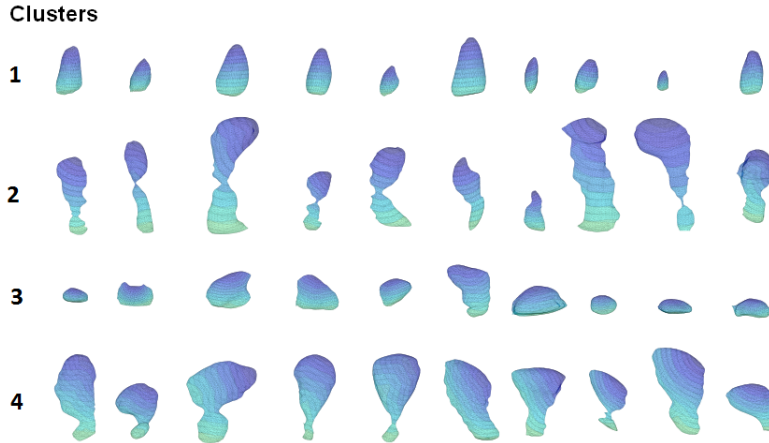


Figura 8.2: Las diez espinas dendríticas más representativas con $p^* \approx 1$ (probabilidad de pertenencia más alta), para las 4 clases. De acuerdo a la clasificación realizada, se pueden buscar coincidencias entre las clases descritas en Peters y Kaiserman-Abramof (1970) mostradas en la Figura 1.3 con las aquí presentadas. Las espinas dendríticas de tipo “Stubby” han sido disgregadas en las clases 1 y 3, las clases “Thin” y “Mushroom” se agrupan bajo la clase 4 mientras que las “Filopodium” coinciden con la clase 2.

8.2. Visualización de las clusters

Para facilitar la visualización de los clusters, las distancias del espacio 7-dimensional así como del espacio 4-dimensional fueron escaladas a un espacio tridimensional de acuerdo al escalamiento multidimensional.

8.2.1. Escalamiento multidimensional

El modelo matemático \mathbf{M} generado a partir de un modelo finito de mixturas devolvió como resultado un espacio n -dimensional para cada espina con la probabilidad de pertenencia para cada una de ellas a cada uno de los clusters. Para conseguir que el resultado del clustering fuera interpretable desde un punto de vista gráfico, la salida del modelo fue proyectada a un espacio tridimensional. Para lograr este cometido, se calculó la divergencia de Jensen-Shannon (Lin, 1991) entre cada par de espinas con el fin de obtener la medida de disimilitud entre las funciones de distribución de probabilidad. La disimilitud entre las distribuciones de probabilidad de las espinas se interpretó como una medida de “distancia”. De esta forma, mediante la aplicación del escalamiento multidimensional (Torgerson, 1952), fue posible mantener las “distancias” entre espinas que se encontraban originalmente en n -dimensiones a un

espacio tridimensional o bidimensional que permite visualizar los resultados.

La divergencia de Jensen-Shannon queda definida como una medida de similitud entre dos distribuciones de probabilidad. Sus principales ventajas sobre otras medidas de divergencia como KL (Kullback y Leibler, 1951) o J (Jeffreys, 1946) radican en su simetría y en que siempre es un valor finito entre los límites 0 y 1.

La divergencia de Jensen-Shannon entre dos distribuciones de probabilidad P y Q se define como

$$JS(P, Q) = \frac{1}{2}KL(P||D) + \frac{1}{2}KL(Q||D),$$

donde $D = \frac{P+Q}{2}$. KL es la divergencia de Kullback-Leibler

$$KL(P||D) = \sum_{c=1}^C p_c \log \frac{p_c}{d_c},$$

donde C denota el número de clusters, p_c la probabilidad de pertenencia a c según una distribución de probabilidad P , y d_c la probabilidad de pertenencia al cluster c bajo la función de probabilidad D .

Después de que la divergencia de Jensen-Shannon haya sido calculada, se obtiene una matriz \mathbf{A} (7342×7342) que representa la “distancia” entre las distribuciones de probabilidad de pertenencia a cada cluster de cada espina. El escalamiento multidimensional (MDS por sus siglas en inglés) se deriva del cálculo $\mathbf{B} = -\frac{1}{2}\mathbf{H} \cdot \mathbf{A} \cdot \mathbf{H}^T$ donde \mathbf{H} es la matriz \mathbf{A} centrada, esto es, $\mathbf{H} = \mathbf{A} - \mathbf{1} \cdot \boldsymbol{\mu}$ donde $\mathbf{1}$ es la matriz con todo unos y $\boldsymbol{\mu}$ es el vector que denota la media de cada columna de \mathbf{A} . Después, a partir de la descomposición por medio de valores propios \mathbf{B} tal que $\mathbf{B} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ se obtienen los vectores propios normalizados (\mathbf{U}) y la matriz diagonal de valores propios $\boldsymbol{\Lambda}$. A continuación, seleccionando los k valores propios más altos para k dimensiones y calculando $\boldsymbol{\Lambda}_k^{\frac{1}{2}}\mathbf{U}_k^T$, donde $\boldsymbol{\Lambda}_k = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$, se obtienen las coordenadas k -dimensionales para las 7342 espinas.

8.2.2. Proyección de las espinas

Las espinas fueron proyectadas a espacios representables visualmente mediante el uso del escalamiento multidimensional para el clustering con 7 clusters con lo que se obtuvo la disposición de las espinas dendríticas de acuerdo a su distancia en un espacio tridimensional. Esta transformación se puede observar en la Figura 8.3.

En la representación se puede observar que la mayoría de los puntos se encuentran centrados alrededor de la media del cluster, con lo que concuerda con los resultados

8. INTERPRETACIÓN DEL MODELO

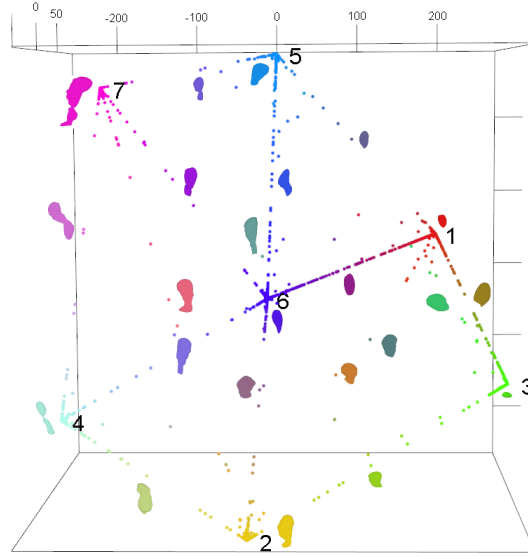


Figura 8.3: Proyección 3D de la distribución 7-dimensional representando la probabilidad de pertenencia de cada espina a cada cluster de acuerdo al escalamiento dimensional. Se muestra una espina con probabilidad $p^* \approx 1$ para cada cluster junto a cada número de cluster. Las espinas fueron coloreadas combinando los colores de los clusters de acuerdo a sus probabilidades de pertenencia a cada cluster. Se mostraron ejemplos de espinas “intermedias” fueron colocados junto a los puntos que las representan.

mostrados en la Tabla 7.1 en la que se indicó que la mayoría de las espinas tenía una probabilidad de pertenencia $p^* > 0,99$. Las espinas “intermedias” cuya p^* se encuentra alejada de 1 fueron coloreadas de acuerdo a su probabilidad de pertenencia a cada cluster lo que dió lugar a un gradiente de colores entre algunos clusters, como por ejemplo entre los clusters 1 y 6, que corroboran la existencia de un continuo de espinas. Además, se puede observar que algunos clusters se encuentran estrechamente relacionados con otros como es el caso del cluster 6, que mantiene espinas intermedias con muchos de los otros clusters. Sin embargo hay otros clusters como por ejemplo el cluster 4 que se encuentra bastante aislado del resto con lo que es fácil de discriminar las espinas de este grupo. Esto coincide con los resultados mostrados en la Tabla 8.1 donde aquellas clases que están muy relacionados necesitan multitud de variables para ser discriminadas del resto mientras que las clases que se encuentran más aisladas son fácilmente identificadas y por lo tanto son necesarias menos variables para reconocerlas.

Del mismo modo se aplicó el escalamiento multidimensional sobre el clustering

de 4 categorías. En este caso se obtuvo una representación bidimensional debido a que, por las propiedades de la matriz de similitud entre espinas, el máximo número de valores propios fue dos. El resultado se muestra en la Figura 8.4.

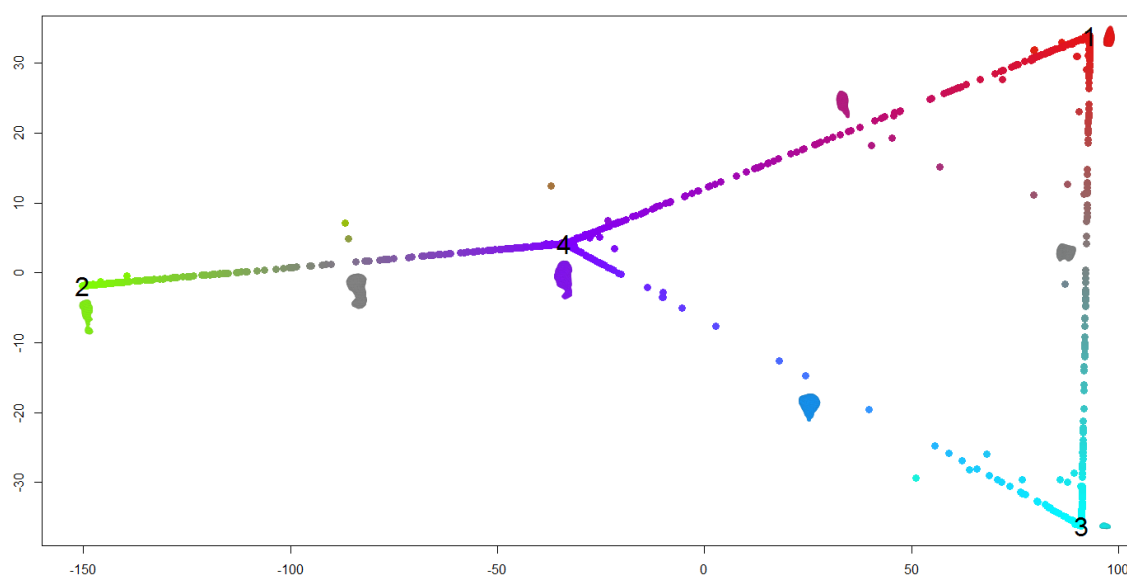


Figura 8.4: Proyección 2D de la distribución 4-dimensional representando la probabilidad de pertenencia de cada espina a cada clase de acuerdo al escalamiento multidimensional. Se muestra una espina con probabilidad $p^* \approx 1$ para cada clase junto a cada número de cluster. Las espinas fueron coloreadas combinando los colores de los clusters de acuerdo a sus probabilidades de pertenencia a cada cluster como en la Figura 8.3. Ejemplos de espinas “intermedias” fueron colocados junto a los puntos que las representan.

En esta ocasión las espinas se encuentran más dispersas que para 7 clusters debido principalmente a que su verosimilitud es inferior. También en este caso el continuo entre tipos de espinas dendríticas es más evidente y, a pesar de ello, la mayoría de espinas se siguen concentrando alrededor del prototipo de cluster.

8.3. Distribución de las espinas por clases en las dendritas

8.3.1. Distribución a partir del compartimento dendrítico, la edad y su combinación

Con el objetivo de obtener una comprensión más profunda, se analizó si la distribución de las clases cambia de forma significativa en función de si varía el compartimento dendrítico, la edad del sujeto o la combinación de ambos (Figura 8.5 B-D). Primero se estudió la distribución de las categorías de acuerdo a su compartimento dendrítico que puede ser observado en la Figura 8.5B. Las clases 2, 4, 6 y 7 presentan una proporción más elevada de espinas dendríticas apicales que la distribución que contempla la población completa (Figura 8.5A). Las clases 1, 3 y 5 se concentran especialmente en las dendritas basales. Para tratar de corroborar la existencia de estas discrepancias se realizó un test de hipótesis χ^2 , esto es, se comprobó si la distribución de las categorías es independiente del compartimento dendrítico (hipótesis nula H_0). El test de hipótesis devolvió un p -valor inferior a 10^{-30} de manera que la hipótesis nula H_0 fue rechazada.

El mismo proceso fue repetido para la edad. La Figura 8.5C muestra como el cluster 6 se encuentra especialmente representada en C40 mientras que el cluster 2 predomina en C85. Para comprobar la independencia entre la distribución de las clases y la edad se realizó de nuevo un test de hipótesis. Los resultados rechazaron la hipótesis nula (el p -valor fue 0,003).

La independencia fue de nuevo testada entre la distribución de las clases y la combinación de compartimentos dendríticos y la edad (Figura 8.5D). La hipótesis nula de nuevo fue rechazada (p -valor $\approx 10^{-35}$) y se concluyó que existe dependencia entre la distribución de las clases y la combinación de compartimento dendrítico y edad.

A pesar de que para todos los casos anteriores la hipótesis nula fue rechazada, un estudio más profundo de la Figura 8.5C muestra que aparentemente las discrepancias en las distribuciones están situadas únicamente en algunas clases y no están repartidas uniformemente. Con el objetivo de diferenciar aquellas categorías que exhibiesen cambios significativos respecto a la distribución global (Figura 8.5A), cada uno de ellos se analizó individualmente. Se realizó un test χ^2 (bondad de ajuste) cluster por cluster de acuerdo a su compartimento dendrítico, edad y la combinación

8.3 Distribución de las espinas por clases en las dendritas

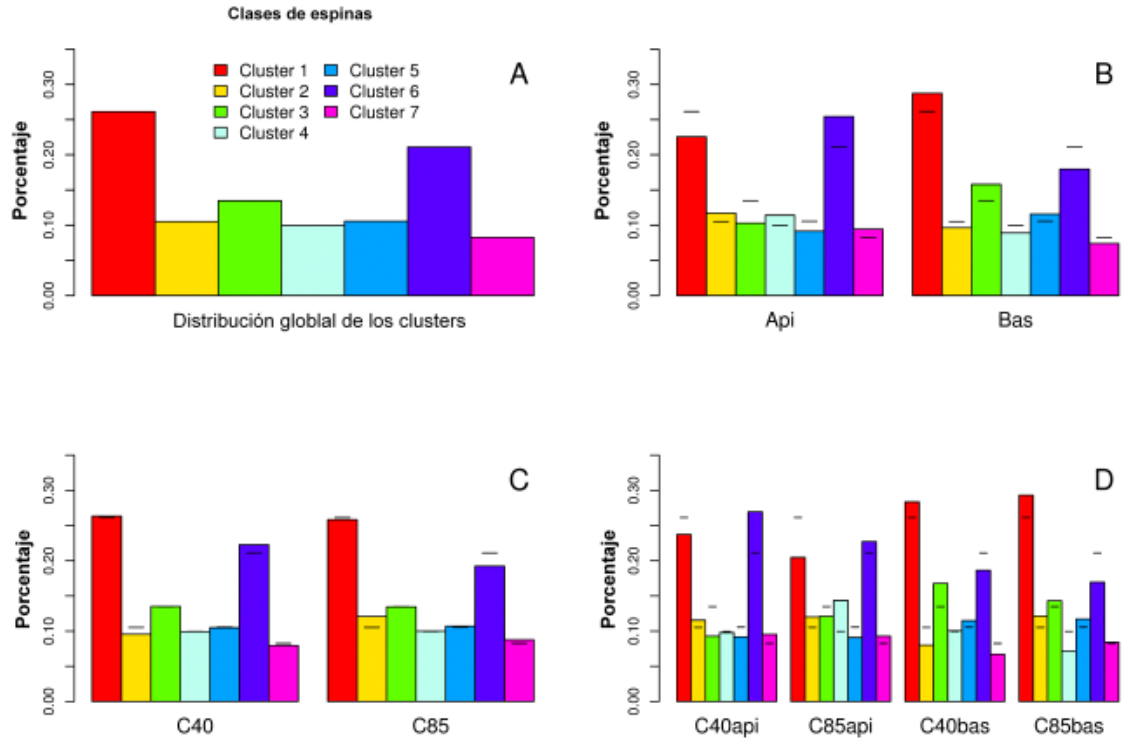


Figura 8.5: Diagrama de barras mostrando la distribución de las espinas que pertenecen a cada uno de los 7 clusters de acuerdo a su máxima probabilidad de pertenencia p^* . (A) Distribución de las espinas en los siete clusters. (B) Distribución de probabilidad de los cluster para las espinas apicales (izquierda) y basales (derecha). (C) Distribución de probabilidad de los clusters para las espinas C40 (izquierda) y C85 (derecha). (D) Distribución de probabilidad de los clusters para la combinación de compartimento dendrítico y edad, C40api (en el extremo izquierdo), C85api (en el centro izquierdo), C40bas (centro derecha), C85bas (en el extremo derecho). Las líneas horizontales en (B), (C) y (D) denotan las alturas mostradas en (A).

de ambos. El resultado del test se muestra en la Tabla 8.3 donde los resultados corroboraron lo que se observó gráficamente, esto es, cuando se trata de encontrar diferencias entre la distribución de las clases de acuerdo a la edad, solamente los clusters 2 y 6 rechazaron la hipótesis nula. Esto quiere decir que son las únicas que variaron significativamente en función de la edad.

Del mismo modo, se estudió la disposición de las espinas sobre el compartimento dendrítico, la edad y la combinación de ambos para el clustering de 4 clases. En esta ocasión, en la Figura 8.6B se puede observar cómo de forma evidente los clusters 2 y 4 sobrepasan para apical el porcentaje de espinas obtenido en la distribución global

8. INTERPRETACIÓN DEL MODELO

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Compartimento dendrítico	***	*	***	**	*	***	*
Edad		*				*	
Combinación	***	***	***	***	*	***	*

Tabla 8.3: Significatividad de los resultados alcanzados con el test de hipótesis entre cada cluster y los compartimentos dendríticos, la edad y la combinación de ambos. El símbolo * simboliza que el p -valor obtenido es inferior al nivel de significatividad 0,05 y que la hipótesis nula es rechazada, ** significa que el p -valor $< 0,001$ y *** indica que el p -valor $< 0,0001$.

mientras que las clusters 1 y 3 quedan notablemente por debajo de la distribución global. Para verificar el resultado observado en el diagrama se llevó a cabo un test de hipótesis χ^2 en el que se comprobó si la distribución de las espinas es independiente del compartimento dendrítico (hipótesis nula H_0). Como resultado se obtuvo un p -valor inferior a 10^{-15} con lo que la hipótesis nula H_0 fue rechazada y se aceptó la hipótesis de que la distribución de las clases depende del compartimento dendrítico.

A continuación se analizó la distribución de las espinas de acuerdo a la edad. Al contrario que en el caso anterior en la Figura 8.6C las barras prácticamente coinciden con la distribución global destacando los casos de las clusters 2 y 4 que se solapan con la línea horizontal que simboliza el valor de la distribución global. El test de hipótesis realizado sobre estas distribuciones arrojó un p -valor $\approx 0,03936$ con lo que, sujeto a un umbral de 0,05, la hipótesis es rechazada pero si el umbral se reduce a, por ejemplo 0,01, no es rechazada. De esta forma la confianza en que la distribución de las espinas sea dependiente de la edad no puede ser verificada.

Por último, se contrasta la distribución de las espinas de acuerdo a la combinación de compartimento dendrítico y edad. Visualmente se aprecia una diferenciación más evidente que para la edad entre la distribución global y la mostrada en la Figura 8.6D. Destaca especialmente el caso de C85api donde el cluster 1 se ve reducida considerablemente en favor del cluster 2. El test de hipótesis obtuvo un p -valor $< 10^{-17}$ por lo que la hipótesis nula fue rechazada y se asumió dependencia entre la distribución de las espinas y la combinación de compartimento dendrítico y edad.

Finalmente, con el fin de localizar qué clusters favorecen los resultados de los test de hipótesis anteriores y profundizar en las dependencias existentes entre clusters y distribución de espinas se repitió el trabajo realizado en el clustering de 7 clases para el clustering con 4 grupos. De nuevo se llevó a cabo un test χ^2 (bondad de ajuste) cluster por cluster de acuerdo a su compartimento dendrítico, edad y la combinación

8.3 Distribución de las espinas por clases en las dendritas

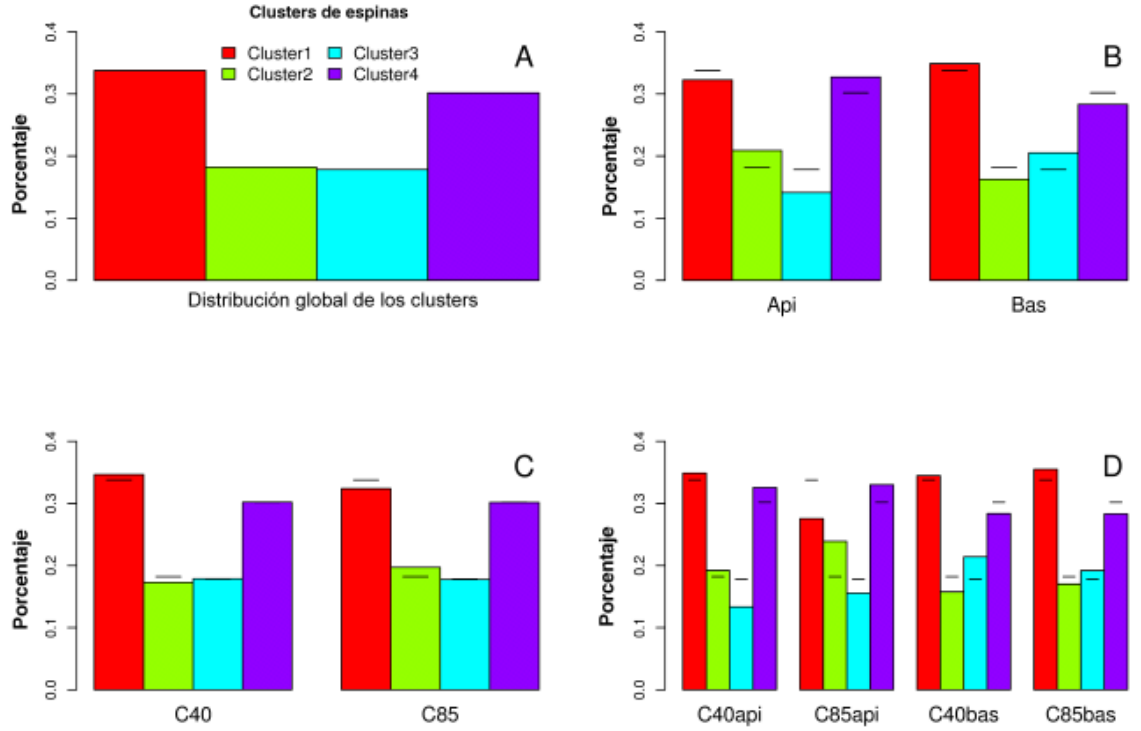


Figura 8.6: Diagrama de barras mostrando la distribución de las espinas que pertenecen a cada uno de los cuatro clusters de acuerdo a la máxima probabilidad de pertenencia p^* . (A) Distribución de las espinas en las cuatro clases. (B) Distribución de probabilidad de los clusters para las espinas apicales (izquierda) y basales (derecha). (C) Distribución de probabilidad de los clusters para las espinas C40 (izquierda) y C85 (derecha). (D) Distribución de probabilidad de los clusters para la combinación de compartimento dendrítico y edad, C40api (en el extremo izquierdo), C85api (en el centro izquierda), C40bas (en el centro derecha) y C85bas (extremo derecha). Las líneas horizontales en (B), (C) y (D) denotan las alturas mostradas en (A).

de ambos. El resultado de aplicar el test se muestra en la Tabla 8.4 que confirma resultados comentados a partir del gráfico, esto es, únicamente la clase 2 es capaz de rechazar la hipótesis para la edad y que para la combinación del compartimento dendrítico y la edad se rechaza H_0 .

8.3.2. Distribución de las espinas a partir de la distancia al soma

El análisis de la distribución de las espinas dendríticas en función de su distancia desde el soma presentado en Benavides-Piccione et al. (2012) fue extendido eva-

8. INTERPRETACIÓN DEL MODELO

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Compartimento dendrítico	*	***	***	**
Edad		*		
Combinación	***	***	***	*

Tabla 8.4: Significatividad de los resultados alcanzados con el test de hipótesis entre cada cluster y los compartimentos dendríticos, la edad y la combinación de ambos. El símbolo * simboliza que el p -valor obtenido es inferior al nivel de significatividad 0,05 y que la hipótesis nula es rechazada, ** significa que el p -valor $< 0,001$ y *** indica que el p -valor $< 0,0001$.

luando la distribución de las clases de acuerdo a su distancia al soma. El número de espinas fue categorizado en secciones de 10 μm de longitud, desde 0 μm (comienzo de la dendrita) hasta 270 μm . Dado que en los rangos desde 0 μm a 40 μm y 230 μm a 270 μm la densidad de espinas fue baja, las espinas en el rango entre 0 μm y 40 μm fueron agrupadas en un único intervalo, al igual que las espinas en el rango 230 μm -270 μm . La distribución de los clusters resultante basada en la distancia desde el soma se muestra en la Figura 8.7. De acuerdo al diagrama de barras, algunos clusters quedan altamente representados en algunas categorías como por ejemplo el cluster 1 cuya proporción se encuentra sobre la probabilidad de la clase global mostrada en la Figura 8.5A entre 0 y 90 μm .

Un test de hipótesis χ^2 fue aplicado para comprobar la dependencia entre las distribuciones de las clases y la distancia al soma. El resultado rechazó la hipótesis nula H_0 (p -valor $\approx 10^{-10}$). Por lo tanto, se asumió dependencia entre la distribución de los clusters y la distancia. Con el objetivo de hallar aquellos clusters que mostrasen cambios significativos respecto a la distribución global de los clusters (Figura 8.5), cada cluster fue evaluado individualmente con un test de hipótesis. Los resultados que se muestran en la Tabla 8.5 indican que los clusters 1, 3 y 4 como aquellos dependientes de la distancia al soma.

Finalmente, el estudio de la distribución de las clases de las espinas dendríticas en función de la distancia fue realizado sobre el clustering de 4 clusters. Al igual que para el clustering de 7 clusters se dispuso de intervalos de 10 μm de longitud, que comenzaron en 0 μm (comienzo de la dendrita) y terminaron en 270 μm . Debido a la baja densidad de espinas dendríticas en las proximidades del soma, las espinas en el intervalo 0 μm -40 μm se agruparon al igual que las que se encuentran entre 230 μm -270 μm . La distribución de los clusters de espinas en función de la distancia

8.3 Distribución de las espinas por clases en las dendritas

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Distancia	*		***	**			

Tabla 8.5: Significatividad de los resultados alcanzados con el test de hipótesis entre cada cluster y la distancia al soma. El símbolo * simboliza que el p -valor obtenido es inferior al nivel de significatividad 0,05 y que la hipótesis nula es rechazada, ** significa que el p -valor $< 0,001$ y *** indica que el p -valor $< 0,0001$.

se muestra en la Figura 8.8. Al igual que ocurriera con los 7 clusters, algunos clusters destacan en ciertos intervalos como es el caso del cluster 1 en los intervalos cercanos al soma. El test de hipótesis para el clustering de 4 clusters basado en distancias obtuvo un p -valor $\approx 7,3 * 10^{-6}$ con lo que H_0 se rechazó y se asumió dependencia entre la distribución de las espinas y su distancia al soma. Por su parte, el resultado del test de hipótesis para cada uno de los clusters se muestra en la Tabla 8.6 en el que se puede ver como en los clusters 1 y 4 no se rechazará la hipótesis de independencia.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Distancia		*	**	

Tabla 8.6: Significatividad de los resultados alcanzados con el test de hipótesis entre cada clase y la distancia al soma. El símbolo * simboliza que el p -valor obtenido es inferior al nivel de significatividad 0,05 y que la hipótesis nula es rechazada, ** significa que el p -valor $< 0,001$ y *** indica que el p -valor $< 0,0001$.

8. INTERPRETACIÓN DEL MODELO

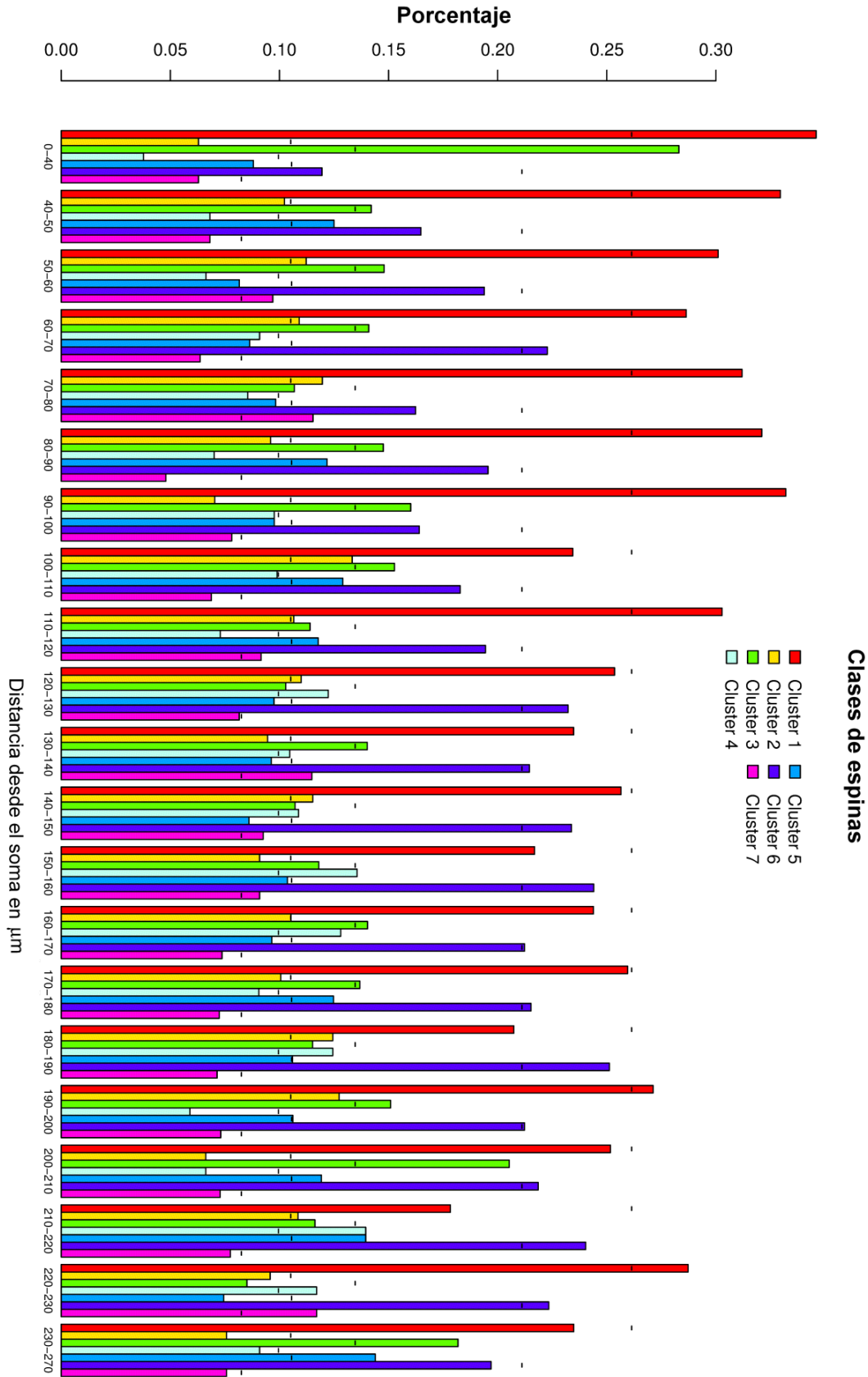


Figura 8.7: Diagrama de barras mostrando la distribución de las espinas que pertenecen a cada una de las 7 clases de acuerdo a la distancia al soma. Las líneas horizontales indican el valor de la distribución para cada una de las clases en la distribución global de las espinas. Algunos intervalos tuvieron que ser agrupados al comienzo (desde 0 a 40 μm) y al final (desde 230 a 270 μm) debido a la baja densidad en estas secciones.

8.3 Distribución de las espinas por clases en las dendritas

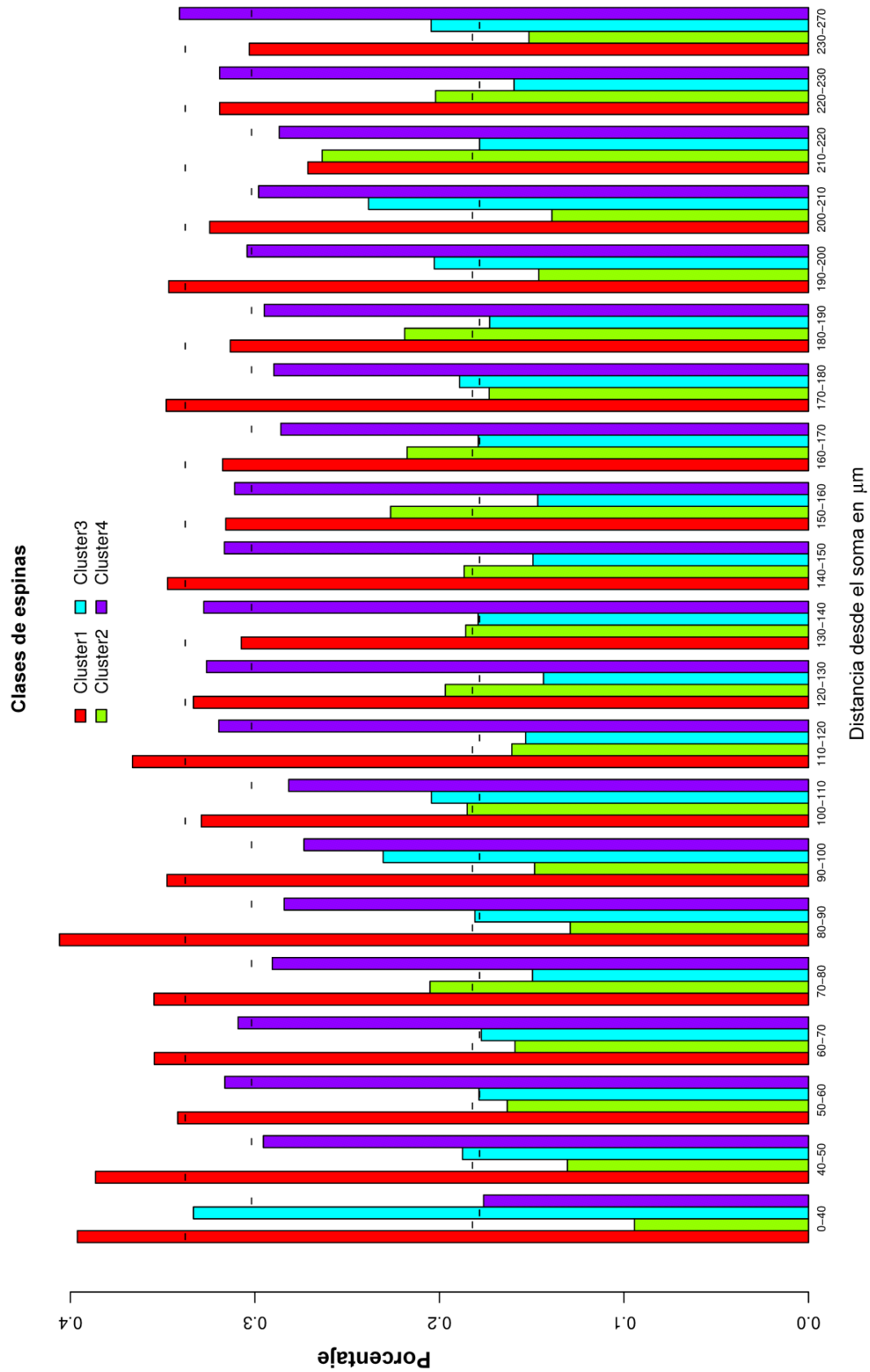


Figura 8.8: Diagrama de barras mostrando la distribución de las espinas que pertenecen a cada una de las 4 clusters de acuerdo a la distancia al soma. Las líneas horizontales indican el valor de la distribución para cada uno de los clusters en la distribución global de las espinas. Algunos intervalos tuvieron que ser agrupados al comienzo (desde 0 a 40 μm) y al final (desde 230 a 270 μm) debido a la baja densidad en los mismos.

8. INTERPRETACIÓN DEL MODELO

9

Conclusiones y trabajo futuro

9.1. Conclusiones

A lo largo del trabajo se han evidenciado distintos aspectos problemáticos en el estudio de los datos en el campo de la neurociencia. Con la realización de este trabajo se han abordado distintas técnicas que han habilitado y automatizado la extracción de conocimiento a partir de los datos de somas y espinas.

9.1.1. Somas

Como se discutió al comenzar el análisis de los somas, la definición del cuerpo neuronal es abstracta y poco precisa. Dado que la morfología del soma depende directamente de su interpretación, resultó imprescindible alcanzar una definición unívoca, objetiva y justificable que se logró mediante la aplicación de métodos matemáticos. Con este enfoque no solo se alcanzó una descripción común sino que además se resolvieron otros aspectos involucrados en este proceso. Algunas de las ventajas que supuso esta técnica son la automatización del proceso, la obtención de cortes menos afilados y más redondeados dado que la herramienta utilizada por los neurocientíficos solo permite realizar cortes rectos, percepción tridimensional del algoritmo de segmentación en contraste con un experto que trabaja sobre una imagen tridimensional representada en un plano con lo que se evita que los cortes varíen en función de la perspectiva desde la que se observa.

A partir de los somas segmentados se investigó el estado del arte con el fin de extraer un conjunto de características definitorias del soma. Las propiedades descritas en el estado del arte no eran adecuadas para la simulación de la morfología de los somas por lo que se introdujo un nuevo conjunto de descriptores basado en Reeb

9. CONCLUSIONES Y TRABAJO FUTURO

graph. De esta forma se logró que con un número reducido de variables, condición necesaria para la obtención de un modelo posterior dado el reducido número de somas, se aproximase la morfología de los somas. Sin embargo este conjunto de propiedades resultaron de difícil interpretación por lo que las características consideradas en el estado del arte se aplicaron como variables emergentes.

El proceso de modelización se desarrolló mediante clustering sobre una red bayesiana. Por las características del problema se definió un nuevo subtipo de red bayesiana denominada dinámica espacial por su semejanza a las redes bayesianas dinámicas pero en este caso aplicadas al espacio. Este enfoque junto a la aplicación de bootstrap permitió reducir el número de arcos y en consecuencia simplificar el modelo. Como resultado del proceso se obtuvieron 3 clusters de somas. A partir del estudio de la estructura de la red y de los descriptores diferenciadores entre clases de somas se pudo observar que las variables angulares fueron prácticamente independientes de las lineales y que las variables angulares fueron la principal fuente de divergencia entre clases. También se pudo comprobar que la distribución de las clases de los somas depende de la localización de la que fueron extraídos.

Finalmente a partir de la parametrización y las relaciones de dependencia del modelo provisto por la red bayesiana dinámica espacial se simularon, por primera vez, nuevos somas. Con ello se cumplió una de los objetivos que se persiguen en proyectos actuales de investigación en los que se pretende alcanzar la simulación del cerebro a nivel molecular dado que se ofrece una técnica para la simulación de uno de sus componentes. De hecho, el método y el conjunto de descriptores propuestos son genéricos y por lo tanto pueden ser extendidos a otros cuerpos tanto en el campo de la neurociencia, como por ejemplo a las espinas, como a otros campos de estudio.

9.1.2. Espinas

Para las espinas se siguió un enfoque distinto al de los somas ya que el objetivo que se persiguió no fue el de simular espinas sino el de distinguir clusters de espinas. Además, con el desarrollo de este apartado del trabajo también se trató de discutir la existencia de un continuo de morfologías de espinas así como la distribución de las clases de espinas en función del compartimento dendrítico y la edad de las personas a lo que hay que añadir también la distribución de las espinas a lo largo de las dendritas. Por lo tanto en este apartado del trabajo se priorizó disponer de variables significativas para el clustering de espinas sobre la simulación por lo que se agregaron

y eliminaron algunas características respecto al soma. De esta forma se obtuvo un conjunto de variables más sencillas de interpretar.

En esta ocasión para el clustering se utilizó un enfoque basado en mixturas finitas de gaussianas multivariantes. Al evaluar los modelos mediante la puntuación BIC se obtuvieron dos máximos, uno con siete clases y otro con cuatro. El segundo de ellos se correspondía con la clasificación más extendida entre los neurocientíficos de manera que el clustering dio lugar a su descripción matemática. Además, se identificó la existencia de un continuo de espinas dado que la probabilidad de pertenencia de algunas era bastante equilibrada. En contraste, la probabilidad de pertenencia de la mayoría de las espinas fue próxima a 1 por lo que quedaron claramente clasificadas. Por lo tanto se confirmó la existencia de clases de espinas a pesar de la continuidad en su morfología.

Mediante algoritmos de clasificación basados en reglas se identificaron las principales características definitorias de cada clase de espinas con lo que se pudo dar una descripción de cada una de ellas. Además, a partir de la disimilitud entre clases y el escalamiento multidimensional se obtuvo una representación gráfica de la “distancia” entre espinas que demostró visualmente la existencia de un continuo de espinas a la vez que la concentración de espinas alrededor de los prototipos de las clases.

Finalmente, el estudio de la distribución de las clases de las espinas en función del compartimento dendrítico, la edad y la distancia respecto al soma manifestó que los tipos de espinas no se distribuyen de forma equitativa.

9.2. Trabajo futuro

Con el desarrollo del trabajo se han elaborado técnicas y herramientas con las que profundizar en la adquisición de conocimiento acerca de la estructura cerebral. Si bien es cierto que los métodos propuestos permiten el estudio de los somas y las espinas desde un nuevo punto de vista, se espera que añadiendo nuevos procedimientos de análisis o mejorando los ya existentes se logre una nueva percepción. Es por ello que se proponen una serie de líneas de trabajo futuro.

La principal tarea futura a realizar consistirá en modelar las variables angulares de la red bayesiana dinámica espacial mediante distribuciones puramente angulares como son la distribución von Mises o la distribución Wrapper Normal. Esto nuevo

9. CONCLUSIONES Y TRABAJO FUTURO

enfoque se encuentra motivado por las dependencias directas de las variables angulares de la red bayesiana con otras variables angulares. Además, de acuerdo a los estudios de las variables segregadoras tanto para los somas como para las espinas las variables angulares son relevantes en la identificación de algunas clases. Un ejemplo de cómo abordar el problema se presenta en la Figura 9.1.

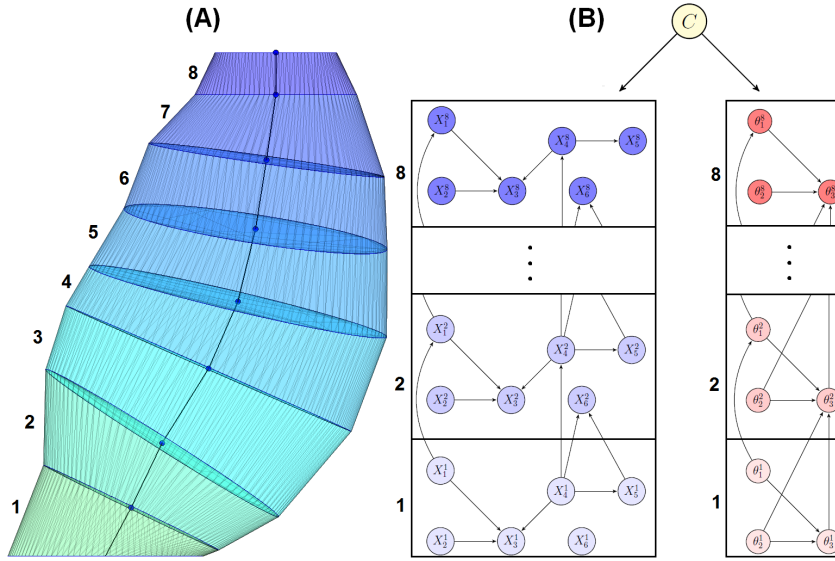


Figura 9.1: Red bayesiana dinámica espacial híbrida: (A) Imagen de una espina con cada una de sus secciones numeradas en orden ascendente. (B) Ejemplo de red bayesiana dinámica espacial angular para la espina de la imagen A. La red bayesiana se divide en dos partes claramente diferenciadas: a la izquierda las variables lineales y la derecha las angulares de manera que, dado el cluster, ambas son independientes. Se puede observar cómo las relaciones de dependencia inter-sección se repiten para cada una de las secciones al igual que las relaciones entre distintas secciones siguen un modelo markoviano de orden 1. En la parte superior se sitúa la variable latente C que categoriza a la espina. Para simplificar la representación no se incluyeron los arcos de dependencia desde C hacia todas las variables.

Los resultados obtenidos para los somas pueden resultar poco significativos dado el reducido número de somas, especialmente al generar el modelo donde el número de variables fue mayor que el número de observaciones. Dado que se espera disponer de nuevos somas en un futuro próximo que sirvan para extender el estudio, se repetirá el análisis completo. De hecho, debido al reducido número de observaciones no se pudieron realizar algunos test de hipótesis con lo que ciertos patrones no pudieron estudiarse en profundidad. Por lo tanto queda pendiente como parte del

trabajo futuro demostrar o rechazar algunas relaciones de dependencia a la espera de más datos.

Durante el procesamiento de los somas algunos tuvieron que ser desestimados debido a que se encontraban demasiado dañados. Esta labor fue realizada manualmente y, para automatizar completamente el procedimiento sin supervisión humana, sería preciso desarrollar algún mecanismo de clasificación que validara los somas.

Otro punto a trabajar consiste en actualizar el algoritmo de orientación de los somas por uno más sofisticado dado que las medidas de los descriptores dependen inmediatamente de la perspectiva desde la que se determinaron. Una opción en este punto podría consistir en utilizar la dendrita apical situándola paralela al eje de cotas. Esta modificación implica desarrollar un método para distinguir las dendritas basales de la dendrita apical.

La distribución de las espinas en función del compartimento dendrítico, la edad, la combinación de ambos y la distancia al soma deben de ser analizados por neurocientíficos con el fin de relacionar la morfología de cada una de los clusters de espinas, información provista por las reglas diferenciadoras entre clusters, con la funcionalidad de las espinas. Con ello se podrían localizar funciones en ciertas parcelas del cerebro o relaciones entre el tipo de actividad cognitiva y la edad del sujeto o la localización de las espinas.

Debido a la relevancia que demuestran tener las espinas en las funciones cognitivas, resulta de gran interés la simulación de estos componentes neuronales. Este objetivo se puede alcanzar extendiendo la técnica aplicada sobre los somas. De esta forma se motiva el uso de la red dinámica espacial angular también para las espinas.

9. CONCLUSIONES Y TRABAJO FUTURO

Apéndices

Resumen de las características de las espinas

A continuación se muestran un conjunto de tablas resumiendo los valores de las características calculadas durante el experimento. Los valores recogidos pueden ser utilizados en el proceso de interpretación de las reglas generadas mediante RIPPER y que son mostradas en el Apéndice 9.2. Cada tabla representa una variable concreta a lo largo de todas las secciones de la espina.

	$ h_1 $	$ h_2 $	$ h_3 $	$ h_4 $	$ h_5 $	$ h_6 $	$ h_7 $	$ h_8 $
Min	0.0015	0.0115	0.0177	0.0246	0.0287	0.0244	0.0111	0.0047
1stQu	0.0479	0.1049	0.1325	0.1454	0.1504	0.1502	0.1391	0.1072
Median	0.0753	0.1529	0.1849	0.1982	0.2036	0.2075	0.1986	0.1610
Mean	0.0846	0.1658	0.1980	0.2129	0.2221	0.2246	0.2147	0.1782
3rdQu	0.1109	0.2108	0.2482	0.2614	0.2703	0.2776	0.2730	0.2299
Max	0.7546	1.0914	0.9469	1.0017	1.0976	1.0792	0.9723	1.0445

Tabla 1: Resumen estadístico de la altura de cada sección.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
Min.	0.0143	0.0359	0.0460	0.0437	0.0366	0.0123	0.0103	0.0059
1st Qu.	0.0362	0.0994	0.1397	0.1580	0.1519	0.1299	0.1088	0.0610
Median	0.0440	0.1148	0.1539	0.1683	0.1679	0.1546	0.1275	0.0709
Mean	0.0470	0.1208	0.1579	0.1676	0.1615	0.1472	0.1248	0.0730
3rd Qu.	0.0546	0.1372	0.1732	0.1783	0.1772	0.1692	0.1430	0.0834
Max.	0.1347	0.2969	0.2825	0.2543	0.2319	0.2202	0.2245	0.1849

Tabla 2: Resumen estadístico del área de cada sección.

. RESUMEN DE LAS CARACTERÍSTICAS DE LAS ESPINAS

	r_2^h	r_3^h	r_4^h	r_5^h	r_6^h	r_7^h	r_8^h
Min.	0.2998	0.3780	0.3118	0.3036	0.2928	0.1675	0.1668
1st Qu.	1.4227	1.575	1.6132	1.5002	1.2726	1.0904	1.0083
Median	1.7147	1.904	2.0253	2.0080	1.8699	1.6686	1.4808
Mean	1.8982	2.018	2.1055	2.0811	1.9752	1.8682	1.6729
3rd Qu.	2.1713	2.338	2.5120	2.5716	2.5173	2.3883	2.1162
Max.	20.2980	9.197	6.3533	6.7537	8.6857	9.6646	10.3850

Tabla 3: Resumen estadístico del achatamiento y elongación entre cada par de secciones.

	e_2	e_3	e_4	e_5	e_6	e_7	e_8
Min.	0.3819	0.3464	0.2215	0.2061	0.1831	0.1731	0.1136
1st Qu.	0.7906	0.6815	0.5962	0.5401	0.5049	0.4866	0.4651
Median	0.8592	0.7768	0.7032	0.6494	0.6170	0.5974	0.5693
Mean	0.8489	0.7677	0.7006	0.6548	0.6261	0.6079	0.5822
3rd Qu.	0.9204	0.8649	0.8107	0.7713	0.7421	0.7238	0.6948
Max.	1.0000	1.0000	0.9992	0.9993	0.9997	0.9984	0.9993

Tabla 4: Resumen estadístico del ratio entre los ejes de cada par de secciones.

	$\cos \alpha_2$	$\cos \alpha_3$	$\cos \alpha_4$	$\cos \alpha_5$	$\cos \alpha_6$	$\cos \alpha_7$	$\cos \alpha_8$
Min.	-0.8245	0.4128	0.3808	0.1660	0.1906	-0.1848	-0.9891
1st Qu.	0.9176	0.9433	0.9474	0.9367	0.9318	0.9221	0.7826
Median	0.9623	0.9742	0.9768	0.9728	0.9714	0.9671	0.9185
Mean	0.9341	0.9567	0.9575	0.9490	0.9449	0.9353	0.8239
3rd Qu.	0.9853	0.9901	0.9911	0.9899	0.9899	0.9880	0.9723
Max.	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tabla 5: Resumen estadístico del coseno del ángulo entre cada par de secciones.

	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
Min.	0.0992	0.1101	0.1602	0.0183	0.1347	0.0790
1st Qu.	1.4255	1.3645	1.3302	1.3042	1.3255	1.3654
Median	1.5721	1.5658	1.5686	1.5795	1.5700	1.5702
Mean	1.5727	1.5673	1.5739	1.5782	1.5695	1.5654
3rd Qu.	1.7191	1.7652	1.8136	1.8502	1.8132	1.7647
Max.	3.0637	3.0633	2.9554	3.0726	3.1042	3.1317

Tabla 6: Resumen estadístico del ángulo colatitud entre cada par de secciones. La primera sección no fue incluida porque es un valor constante.

	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
Min.	0.0019	0.0017	0.0019	0.0015	0.0000	0.0005	0.0010
1st Qu.	1.3576	1.3786	1.3910	1.4405	1.4812	1.5663	1.6926
Median	2.2450	2.1760	2.1936	2.1910	2.3929	3.1875	3.9511
Mean	2.7316	2.7103	2.7409	2.7644	2.8743	3.1578	3.3902
3rd Qu.	4.2519	4.2523	4.3371	4.3322	4.5079	4.7224	4.7897
Max.	6.2824	6.2812	6.2812	6.2829	6.2817	6.2831	6.2822

Tabla 7: Resumen estadístico del ángulo azimutal entre cada par de secciones.

. RESUMEN DE LAS CARACTERÍSTICAS DE LAS ESPINAS

Reglas generadas con RIPPER

Reglas para el cluster de 7 categorías

Cluster 1

$$(|h_7| \leq 0,1791) \ \& \ (r_3^h \leq 2,7439) \ \& \ (r_6^h \geq 1,9660) \rightarrow class = 1$$

$$(r_7^h \geq 1,5084) \ \& \ (|h_7| \leq 0,2086) \ \& \ (r_4^h \leq 3,0194) \ \& \ (e_5 \geq 0,6287) \ \& \ (r_5^h \geq 1,8055) \rightarrow class = 1$$

$$(|h_7| \leq 0,2010) \ \& \ (r_6^h \geq 1,6653) \ \& \ (r_4^h \leq 3,1141) \ \& \ (\cos \alpha_6 \geq 0,9450) \ \& \ (\cos \alpha_7 \geq 0,9593) \rightarrow class = 1$$

$$(|h_7| \leq 0,2089) \ \& \ (r_6^h \geq 1,7844) \ \& \ (r_4^h \leq 3,1148) \ \& \ (\cos \alpha_6 \geq 0,9439) \ \& \ (\cos \alpha_5 \geq 0,9564) \rightarrow class = 1$$

Cluster 2

$$(\cos \alpha_6 \leq 0,9408) \ \& \ (\cos \alpha_6 \leq 0,8136) \rightarrow class = 2$$

$$(a_4 \geq 0,1857) \ \& \ (r_2^h \geq 2,1276) \rightarrow class = 2$$

$$(\cos \alpha_7 \leq 0,9471) \ \& \ (|h_4| \geq 0,1714) \ \& \ (\cos \alpha_7 \leq 0,8547) \rightarrow class = 2$$

$$(\cos \alpha_5 \leq 0,9042) \rightarrow class = 2$$

Cluster 3

$$(a_1 \leq 0,03512) \rightarrow class = 3$$

Cluster 4

$$(r_5^h \leq 1,2665) \rightarrow class = 4$$

$$(r_5^h \leq 1,5633) \ \& \ (e_3 \leq 0,7370) \rightarrow class = 4$$

Cluster 5

$$(\phi_6 \leq 2,0779) \ \& \ (\phi_5 \leq 2,2250) \ \& \ (\phi_4 \leq 2,1371) \rightarrow class = 5$$

$$(\phi_6 \leq 3,4115) \ \& \ (\phi_4 \leq 3,3580) \ \& \ (\phi_5 \leq 3,4291) \ \& \ (|h_6| \geq 0,1751) \rightarrow class = 5$$

Cluster 6

$(a_1 \geq 0,0432) \ \& \ (|h_6| \leq 0,2786) \ \& \ (r_4^h \leq 1,8025) \ \& \ (\cos \alpha_4 \geq 0,9227) \rightarrow class = 6$

$(r_4^h \leq 2,1621) \ \& \ (r_6^h \geq 1,2498) \ \& \ (a_1 \geq 0,0450) \ \& \ (e_6 \geq 0,5641) \rightarrow class = 6$

$(a_1 \geq 0,0406) \ \& \ (|h_5| \leq 0,2965) \ \& \ (a_5 \leq 0,1681) \ \& \ (a_1 \leq 0,0678) \rightarrow class = 6$

$(a_1 \geq 0,0418) \ \& \ (|h_6| \leq 0,2815) \ \& \ (|h_2| \geq 0,1566) \ \& \ (\cos \alpha_4 \geq 0,9348) \rightarrow class = 6$

Cluster 7

$(r_6^h \leq 1,3971) \ \& \ (\phi_5 \leq 2,1000) \ \& \ (\phi_4 \leq 3,3981) \rightarrow class = 7$

$(r_6^h \leq 1,3971) \ \& \ (\phi_4 \leq 2,8734) \ \& \ (|h_5| \geq 0,2886) \ \& \ (a_1 \leq 0,0691) \rightarrow class = 7$

$(r_6^h \leq 1,5705) \ \& \ (\phi_5 \leq 2,8734) \ \& \ (\phi_4 \leq 3,3409) \ \& \ (|h_5| \geq 0,2589) \rightarrow class = 7$

$(r_6^h \leq 1,6531) \ \& \ (\phi_5 \leq 2,8198) \ \& \ (\phi_6 \geq 2,9387) \ \& \ (a_4 \geq 0,1455) \rightarrow class = 7$

Reglas para el cluster de 4 categorías

Cluster 1

$$\begin{aligned} &(|h_7| \leq 0,2096) \ \& \ (r_4^h \leq 2,7579) \ \& \ (e_4 \geq 0,6726) \ \& \ (\cos \alpha_5 \geq 0,9563) \rightarrow class = 1 \\ &(|h_7| \leq 0,2010) \ \& \ (r_4^h \leq 2,7650) \ \& \ (\cos \alpha_5 \geq 0,9479) \ \& \ (a_6 \geq 0,1404) \ \& \ (\cos \alpha_6 \geq 0,9757) \rightarrow class = 1 \\ &(|h_7| \leq 0,2229) \ \& \ (a_1 \geq 0,0329) \ \& \ (r_6^h \geq 1,6908) \ \& \ (\cos \alpha_5 \geq 0,9587) \ \& \ (\cos \alpha_6 \geq 0,9229) \rightarrow class = 1 \\ &(|h_7| \leq 0,2337) \ \& \ (r_4^h \leq 2,9868) \ \& \ (r_6^h \geq 1,4670) \ \& \ (\cos \alpha_7 \geq 0,9316) \rightarrow class = 1 \end{aligned}$$

Cluster 2

$$\begin{aligned} &(r_5^h \leq 1,5125 \ \& \ (r_5^h \leq 1,1124) \rightarrow class = 2 \\ &(r_5^h \leq 1,5821 \ \& \ (e_4 \leq 0,6885) \rightarrow class = 2 \\ &(r_5^h \leq 1,7053 \ \& \ (e_3 \leq 0,6774) \rightarrow class = 2 \\ &(r_6^h \leq 1,2917) \rightarrow class = 2 \end{aligned}$$

Cluster 3

$$(a_1 \leq 0,0351) \rightarrow class = 3$$

Cluster 4

$$\begin{aligned} &(r_7^h \leq 1,7988) \ \& \ (r_5^h \geq 1,2297) \ \& \ (|h_7| \geq 0,2137) \rightarrow class = 4 \\ &(|h_7| \geq 0,1885) \ \& \ (r_5^h \geq 1,1445) \ \& \ (r_7^h \leq 1,6078) \rightarrow class = 4 \\ &(r_7^h \leq 2,1106) \ \& \ (r_5^h \geq 1,0719) \ \& \ (|h_6| \geq 0,1894) \rightarrow class = 4 \end{aligned}$$

. REGLAS GENERADAS CON RIPPER

Bibliografía

- Azadeh Alavi, Brenton Cavanagh, Gervase Tuxworth, Adrian Meedeniya, Alan Mackay-Sim, y Michael Blumenstein. Automated classification of dopaminergic neurons in the rodent brain. In *International Joint Conference on Neural Networks (IJCNN 2009)*, pages 81–88. IEEE, 2009. 24, 25, 26
- Luis Álvarez, Luis Baumela, Pedro Henríquez, y Pablo Márquez-Neila. Morphological snakes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pages 2197–2202. IEEE, 2010. 68
- Miguel Baguear. Morphological study of dendritic spines. ETSI Informática, UPM, 2011. 65, 67, 68
- Jonathan M Bélisle, Leonard A Levin, y Santiago Costantino. High-content neurite development study using optically patterned substrates. *PLOS ONE*, 7(4):e35911, 2012. 25
- Ruth Benavides-Piccione, Farid Hamzei-Sichani, Inmaculada Ballesteros-Yáñez, Javier DeFelipe, y Rafael Yuste. Dendritic size of pyramidal neurons differs among mouse cortical regions. *Cerebral Cortex*, 16(7):990–1001, 2006. 25
- Ruth Benavides-Piccione, Isabel Fernaud-Espinosa, Victor Robles, Rafael Yuste, y Javier DeFelipe. Age-based comparison of human dendritic spine structure using complete three-dimensional reconstructions. *Cerebral Cortex*, 23(8):1798–1810, 2012. 66, 67, 93
- Tobias Bonhoeffer y Rafael Yuste. Spine motility: phenomenology, mechanisms, and function. *Neuron*, 35(6):1019–1027, 2002. 65
- Juan P Brito, Susana Mata, Sofia Bayona, Luis Pastor, Javier DeFelipe, y Ruth Benavides-Piccione. Neuronize: A tool for building realistic neuronal cell morphologies. *Frontiers in Neuroanatomy*, 7, 2013. 12
- Zhang Cha y Chen Tsuhan. Efficient feature extraction for 2d/3d objects in mesh representation. In *International Conference on Image Processing*, volume 3, pages 935–938 vol.3, 2001. 24

BIBLIOGRAFÍA

- Paolo Cignoni. MeshLab. URL <http://meshlab.sourceforge.net/>. 12, 55
- William W. Cohen. Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning*, pages 115–123. Morgan Kaufmann, 1995. 81
- Heron de Alejandría. *Metrica*. 60. 24
- Arthur P Dempster, Nan M Laird, y Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977. 41, 76
- Daniella Dumitriu, Rosa Cossart, Josh Huang, y Rafael Yuste. Correlation between axonal morphologies and synaptic input kinetics of interneurons from mouse visual cortex. *Cerebral Cortex*, 17(1):81–91, 2007. 25
- Anna Dunaevsky, Ayumu Tashiro, Ania Majewska, Carol Mason, y Rafael Yuste. Developmental regulation of spine motility in the mammalian central nervous system. *Proceedings of The National Academy of Sciences*, 96(23):13438–13443, 1999. 66
- Bradley Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, pages 1–26, 1979. 38
- Bradley Efron y Robert J Tibshirani. *An Introduction to the Bootstrap*. CRC Press, 1994. 38
- Eduardo Ceretta Dalla Favera y Waldemar Celes. Ambient occlusion using cone tracing with scene voxelization. In *25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 142–149. IEEE, 2012. 13
- Usama Fayyad, Gregory Piatetsky-Shapiro, y Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37, 1996. 7, 8, 65
- Chris Fraley y Adrian E. Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal Of The American Statistical Association*, 97:611–631, 2002. 75, 77
- Nir Friedman, Moises Goldszmidt, y Abraham Wyner. Data analysis with bayesian networks: A bootstrap approach. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 196–205. Morgan Kaufmann Publishers Inc., 1999. 38
- Johannes Fürnkranz. Pruning algorithms for rule learning. *Machine Learning*, 27(2):139–172, 1997. 82

- Radim Halír y Jan Flusser. Numerically stable direct least squares fitting of ellipses. In *Proceedings 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG*, volume 98, pages 125–132, 1998. 30
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, y Ian H. Witten. The weka data mining software: An update. *SIGKDD Explorations Newsletters*, 11(1):10–18. 81
- Bruce Harland. Recovery of function after lesions of the anterior thalamic nuclei: Cal neuromorphology, 2013. 24, 25
- Max Henrion. Propagating uncertainty in bayesian networks by probabilistic logic sampling. In *UAI*, pages 149–164, 1988. 52
- Armando Herrera-Arellano, Judith Miranda-Sánchez, Pedro Ávila-Castro, Sara Herrera-Álvarez, Jesús Enrique Jiménez-Ferrer, Alejandro Zamilpa, Rubén Román-Ramos, Héctor Ponce-Monter, y Jaime Tortoriello. Clinical effects produced by a standardized herbal medicinal product of hibiscus sabdariffa on patients with hypertension. a randomized, double-blind, lisinopril-controlled clinical trial. *Planta Medica*, 73(01):6–12, 2007. 5, 66, 75, 78
- Masaki Hilaga, Yoshihisa Shinagawa, Taku Kohmura, y Toshiyasu L Kunii. Topology matching for fully automatic similarity estimation of 3d shapes. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 203–212. ACM, 2001. 27, 68
- Bob Jacobs, Lori Driscoll, y Matthew Schall. Life-span dendritic and spine changes in areas 10 and 18 of human cortex: A quantitative golgi study. *Journal of Comparative Neurology*, 386(4):661–680, 1997. 6
- Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical And Physical Sciences*, 186(1007):453–461, 1946. 87
- Haruo Kasai, Masahiro Fukuda, Satoshi Watanabe, Akiko Hayashi-Takagi, y Jun Noguchi. Structural dynamics of dendritic spines in memory and cognition. *Trends in Neurosciences*, 33(3):121–129, 2010. 66
- Michael Kazhdan, Matthew Bolitho, y Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, 2006. 15, 16

BIBLIOGRAFÍA

- Cleopatra Kozlowski y Robby M Weimer. An automated method to quantify microglia morphology and application to monitor activation state longitudinally in vivo. *PLOS ONE*, 7(2):e31814, 2012. 25
- Solomon Kullback y Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, pages 79–86, 1951. 87
- Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. 86
- M Masseroli, A Bollea, y G Forloni. Quantitative morphology and shape classification of neurons by computerized image analysis. *Computer Methods and Programs in Biomedicine*, 41(2):89–99, 1993. 24, 25, 26
- Hajime Matsuzaki, Halina Loi, Shoulian Dong, Ya-Yu Tsai, Joy Fang, Jane Law, Xiaojun Di, Wei-Min Liu, Geoffrey Yang, y Guoying Liu. Parallel genotyping of over 10,000 snps using a one-primer assay on a high-density oligonucleotide array. *Genome Research*, 14(3):414–425, 2004. 65
- Andrew Matus. Actin-based plasticity in dendritic spines. *Science*, 290(5492):754–758, 2000. 66
- Geoffrey J McLachlan y Kaye E Basford. Mixture models. inference and applications to clustering. *Statistics: Textbooks and Monographs*, 1, 1988. 14, 76
- Erik Meijering. Neuron tracing in perspective. *Cytometry Part A*, 77(7):693–704, 2010. 24, 25, 26
- Joseph SB Mitchell, David M Mount, y Christos H Papadimitriou. The discrete geodesic problem. *SIAM Journal on Computing*, 16(4):647–668, 1987. 68
- Radhakrishnan Nagarajan, Marco Scutari, y Sophie Lebre. *Bayesian Networks in R with Applications in Systems Biology*. Springer, 2013. 37
- Esther A Nimchinsky, Bernardo L Sabatini, y Karel Svoboda. Structure and function of dendritic spines. *Annual Review of Physiology*, 64(1):313–353, 2002. 4
- Zoltan Nusser, David Naylor, y Istvan Mody. Synapse-specific contribution of the variation of transmitter concentration to the decay of inhibitory postsynaptic currents. *Biophysical Journal*, 80(3):1251–1261, 2001. 65
- Joseph O’Rourke. Computational geometry column 35. *ACM SIGACT News*, 30(2):31–32, 1999. 68

- Sandeep Patil y B Ravi. Voxel-based representation, display and thickness analysis of intricate shapes. In *Ninth International Conference on Computer Aided Design and Computer Graphics*, pages 6–pp. IEEE, 2005. 67
- Alan Peters y Ita R Kaiserman-Abramof. The small pyramidal neuron of the rat cerebral cortex. the perikaryon, dendrites and spines. *American Journal of Anatomy*, 127(4): 321–355, 1970. III, IV, 5, 7, 66, 78, 86
- Tingwei Quan, Ting Zheng, Zhongqing Yang, Wenxiang Ding, Shiwei Li, Jing Li, Hang Zhou, Qingming Luo, Hui Gong, y Shaoqun Zeng. Neurogps: Automated localization of neurons for brain circuits using l1 minimization model. *Scientific Reports*, 3, 2013. 27
- J. R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5:239–266, 1990. 82
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL <http://www.R-project.org>. 75
- Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2): 461–464, 1978. 44
- Marco Scutari. Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35(3):1–22, 2010. URL <http://www.jstatsoft.org/v35/i03/>. 37
- Lior Shapira, Ariel Shamir, y Daniel Cohen-Or. Consistent mesh partitioning and skeletonisation using the shape diameter function. *The Visual Computer*, 24(4):249–259, 2008. 17
- Benjamin A Smith, Hugo Roy, Paul De Koninck, Peter Grütter, y Yves De Koninck. Dendritic spine viscoelasticity and soft-glassy nature: Balancing dynamic remodeling with structural stability. *Biophysical Journal*, 92(4):1419–1430, 2007. 5
- Nelson Spruston. Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3):206–221, 2008. 3, 5, 6
- András Stelescu, János Sümegi, Ildikó Wéber, András Birinyi, y Ervin Wolf. Somato-dendritic morphology and dendritic signal transfer properties differentiate between fore- y hindlimb innervating motoneurons in the frog rana esculenta. *BMC Neuroscience*, 13(1):68, 2012. 27

BIBLIOGRAFÍA

- Joan Stiles y Terry L Jernigan. The basics of brain development. *Neuropsychology Review*, 20(4):327–348, 2010. 2
- Johan WH Tangelder y Remco C Veltkamp. A survey of content based 3d shape retrieval methods. *Multimedia Tools and Applications*, 39(3):441–471, 2008. 68
- Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952. 86
- Arete Tsiola, Farid Hamzei-Sichani, Zita Peterlin, y Rafael Yuste. Quantitative morphologic classification of layer 5 neurons from mouse primary visual cortex. *Journal of Comparative Neurology*, 461(4):415–428, 2003. 25
- Thomas Willis. *Cerebri Anatome: Cui Accessit Nervorum Descriptio et Usus*. Apud Gerbrandum Schagen, 1663. 1
- Rafael Yuste y Winfried Denk. Dendritic spines as basic functional units of neuronal integration. *Nature*, 375(6533):682–684, 1995. 65
- Rafael Yuste, Ania Majewska, y Knut Holthoff. From form to function: Calcium compartmentalization in dendritic spines. *Nature Neuroscience*, 3(7):653–659, 2000. 65
- Ying Zhang, Yuanming Feng, Calvin R Justus, Wenhuan Jiang, Zhigang Li, Jun Q Lu, R Scott Brock, Matthew K McPeck, Douglas A Weidner, y Li V Yang. Comparative study of 3d morphology and functions on genetically engineered mouse melanoma cells. *Integrative Biology*, 4(11):1428–1436, 2012. 25
- Carl Zimmer y Robert Clark. Secrets of the brain. *National Geographic*, 2014. 1